McGRAW-HILL
INDUSTRIAL ORGANIZATION AND MANAGEMENT SERIES
L. C. MORROW, *Consulting Editor*
* * *

PRINCIPLES OF PERSONNEL TESTING

McGRAW-HILL
INDUSTRIAL ORGANIZATION AND MANAGEMENT SERIES
L. C. Morrow, *Consulting Editor*
* * *

PRINCIPLES OF PERSONNEL TESTING

McGRAW-HILL INDUSTRIAL ORGANIZATION AND MANAGEMENT SERIES

L. C. Morrow, *Consulting Editor*

★ ★ ★

Bethel, Atwater, Smith, and Stackman—*Industrial Organization and Management*

Bethel, Atwater, Smith, and Stackman—*Essentials of Industrial Management*

Carroll—*How to Chart Timestudy Data*

Carroll—*How to Control Production Costs*

Feigenbaum—*Quality Control: Principles, Practice, and Administration*

Finlay, Sartain, and Tate—*Human Behavior in Industry*

Gardner—*Profit Management and Control*

Grant—*Statistical Quality Control*

Hannaford—*Conference Leadership in Business and Industry*

Heide—*Industrial Process Control by Statistical Methods*

Hyde—*Fundamentals of Successful Manufacturing*

Immer—*Layout Planning Techniques*

Juran—*Quality-control Handbook*

Kalsem—*Practical Supervision*

Kimball and Kimball—*Principles of Industrial Organization*

Landy—*Production Planning and Control*

Lawshe—*Principles of Personnel Testing*

Maynard, Stegemerten, and Schwab—*Methods-Time Measurement*

Michael—*Wage and Salary Fundamentals and Procedures*

Neuschel—*Streamlining Business Procedures*

Smyth and Murphy—*Job Evaluation and Employee Rating*

Staniar—*Plant Engineering Handbook*

Tootle—*Employees Are People*

Young—*Personnel Manual for Executives*

PRINCIPLES OF PERSONNEL TESTING

by

C. H. LAWSHE, Jr.

Professor of Psychology, Purdue University

McGRAW-HILL BOOK COMPANY, INC.

NEW YORK TORONTO LONDON

1948

PRINCIPLES OF PERSONNEL TESTING

VIII

To
MURIEL

PREFACE

The recent war years have clearly demonstrated the effectiveness of personnel tests both in industry and in the military services. That they are genuinely useful managerial tools is attested by the fact that psychological tests have been employed in some industries for almost two decades. However, the adoption of personnel tests in business and industry has been retarded somewhat by the lack of trained personnel to administer testing programs and by the lack of information about tests on the part of those in managerial capacities.

It is the hope of the author that this book will prove useful to those now in or soon to be in managerial positions as a statement of what can legitimately be expected of tests and as a guide to the establishment of the policy framework within which a testing program must function. It is further hoped that this book will be useful in the training of those who will eventually administer testing programs. As every applied psychologist knows, it is often too great a step from the theory and perfectionism of the classroom to the reality and pragmatism of business and industry. Because of this fact, emphasis has been placed upon procedure rather than theory and upon results rather than rationale.

Acknowledgments are justly due those who contributed directly and indirectly, especially to the author's colleague and former teacher, Dr. Joseph Tiffin, who with the author inaugurated the Purdue Industrial Personnel Testing Institute and who collaborated on many of the research studies reported here; to Dr. F. B. Knight, the author's immediate superior at the university, whose generous research policy has made possible many of the author's investigations; and to A. C. Eckerman for wise counsel and permission to adapt material prepared by him for Chap. XIV and Appendix B. The author is indebted to Prof. Ronald

A. Fisher and Dr. Frank Yates, also to Oliver & Boyd Ltd., Edinburgh, for permission to abridge Table IX from their book, *Statistical Tables for Biological, Agricultural, and Medical Research.* Further acknowledgment is made of the kind assistance of Max H. Forster, William H. Angoff, R. A. Sherman, D. E. Cole, and Dr. Frank Stump, each of whom read critically all or part of the manuscript. Acknowledgment is made anonymously of the help of many persons in the industrial field with whom the author has had consulting relationships, and whose questions and problems have contributed toward the crystallization of the author's point of view here presented. To all of these, to the many other friends too numerous to mention, and to the author's secretary, Mrs. Elaine Bonnet, who processed the manuscript through its several stages, sincere thanks and appreciation are extended.

C. H. Lawshe, Jr.

West Lafayette, Ind.
February, 1948

CONTENTS

CHAPTER I

THE BASIS OF PERSONNEL TESTING

Human beings differ. They differ in their physical attributes, their abilities, their temperaments, their interests, and their attitudes. Because they differ in these personality characteristics, they naturally differ in the ways in which they perform their jobs. Some employees in a given group are better than others in spite of how good or how poor the group is as a whole. How great are these differences; how are they distributed; and what has personnel testing to contribute to the employment or upgrading situation in the light of these facts? The purpose of this opening chapter is to deal specifically with these questions.

Production Output.—One company engaged in the manufacturing of diamond dies which are used for extruding fine tungsten wire had eleven men on the job of drilling. These men were engaged in the process of drilling holes through industrial diamonds. They were all selected because they showed promise of succeeding in that work. The obvious misfits, the floaters and ne'er-do-wells, and the disinterested were never hired. The eleven in question were chosen because, in terms of the usual employment procedures, they seemed to be good prospects. A later analysis of their production records, however, revealed important differences in their output. The table on the following page shows the number of dies that each drilled in a six week's period. Note that one individual processed 741 dies while another processed only 288. The group average was approximately 450; and if this is considered 100 per cent, the best man produced 165 per cent, or 65 per cent *above* average, and the poorest man produced 64 per cent, or 36 per cent *below* average. Stated another way, the best employee produced about 2.6 times as much as the poorest man in the group. It should be remembered that

these eleven individuals were considered equally good risks at the time of employment.

Employee A	741 dies
Employee B	572 dies
Employee C	519 dies
Employee D	498 dies
Employee E	456 dies
Employee F	436 dies
Employee G	419 dies
Employee H	372 dies
Employee I	342 dies
Employee J	304 dies
Employee K	288 dies

Quality of Work.—Employees differ not only in their rate of production but also in the quality of their work. Forty-five employees in the toolroom of a company engaged in making a particular precision part for aircraft engines were each given

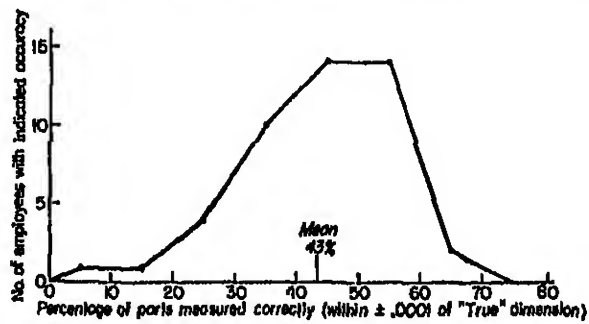


FIG. 1-1.—Distribution of accuracy scores made by forty-five toolroom employees in measuring nineteen pieces of metal with vernier micrometers.

nineteen metal pieces of varying sizes and shapes and asked to measure them with standard vernier micrometers.[1] Each part had been previously measured by ultraprecision methods, and each employee was given an accuracy score that was the percentage of pieces he measured correctly within 0.0001 inch of the "true" measurement. Figure 1-1 shows the distribution of these

[1] LAWSHE, C. H., JR., AND TIFFIN, JOSEPH. The accuracy of precision instrument measurement in industrial inspection. *J. appl. Psychol.*, 1945, 29, 413-419.

scores. Note that one man had an accuracy score between zero and 10 per cent, one other had a score between 10 and 20 per cent. Note also that two men had scores between 60 and 70 per cent and that the mean or average of the group was 43 per cent. The original data from which the figure was prepared indicate that the poorest performer had measured 5 per cent (actually only one) of the parts correctly while the best performer meas-

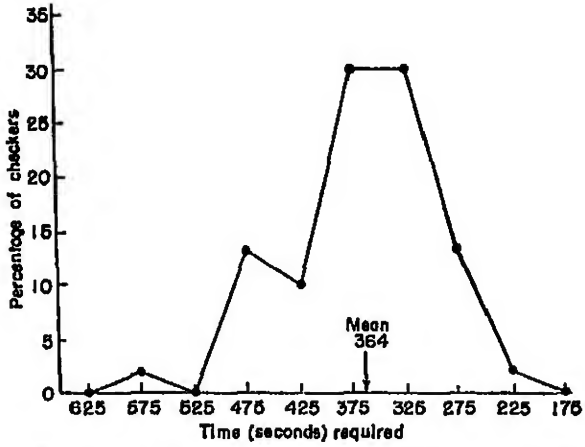


FIG. 1-2.—Distribution of times required by forty-six order checkers to check a standard order of groceries.

ured 63 per cent, or twelve, of the nineteen pieces correctly. Note that the curve approaches the shape of a bell.

Ways of Evaluating Job Performance.—There are many ways of evaluating job performance which will be treated systematically and more thoroughly in Chap. III. For the time being it is sufficient to point out that, when production data are used for this evaluation, the figures always reflect quantity, quality, or a combination of the two. Just what form the facts assume is quite often a function of the nature of the job or the work. A few examples from a variety of jobs will suffice.

Grocery Checkers.—In one study involving cashier-checkers in a supermarket, forty-six employees checked a standard customer order and were given time scores. Figure 1-2 shows that the time in seconds required for checking this particular order

ranged from a low of 225 seconds to a high of 575 seconds with a mean or average of 364. Checking, like many jobs, is one in which skill and speed are developed with experience. An analysis of the records indicated that seventeen of the forty-six checkers had been on the job less than six months. Another curve (Fig. 1-3) was prepared in which only the twenty-nine employees with six months' experience or more were included. Note,

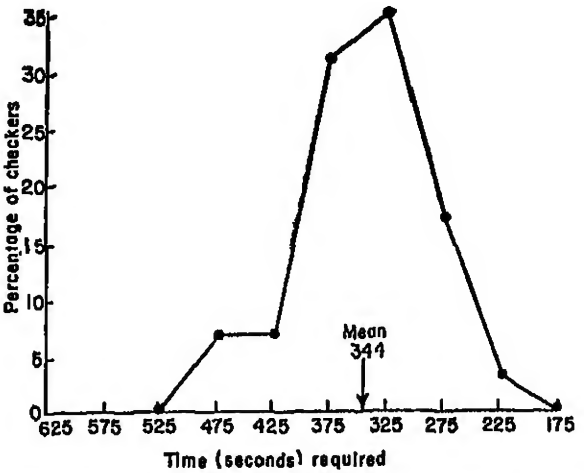


FIG. 1-3.—Distributions of times required by twenty-nine experienced order checkers to check a standard order of groceries.

however, that the time range, although not so large, is still from 225 to 475 seconds and the average is 344. Even when the less experienced checkers or the learners are eliminated, the best performers are 35 per cent above average and the poorest are 53 per cent below average. The best ones require a little less than half the time required by the slower ones, *even when the beginners are not included.* The implications in terms of number of checkers needed to handle a particular volume during rush hours seem clear.

Wool Pullers.—In packing houses, employees are engaged to pull wool from sheep pelts by hand. One study of number of pelts pulled involved thirteen employees, and the results are shown in Fig. 1-4. During the time studied, the average number

of pelts pulled was 156. One man, however, pulled 215, or 138 per cent of average, and another pulled only 119, or 76 per cent of average. There again, the ratio of the best to the poorest is almost two to one.

Department-store Salespersons.—In the sales field, adequacy of sales personnel is frequently measured in terms of volume. In the curtain and drapery department of one large department

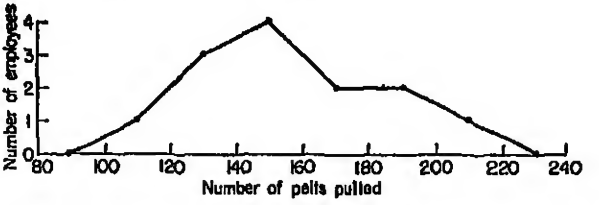


FIG. 1-4.—Distribution of number of pelts pulled by thirteen packing-house employees.

store, eighteen clerks averaged a little over $4,000 each in one monthly period. One employee, however, sold over $7,100 worth of merchandise (177 per cent of average), and another sold only about $500 worth (13 per cent of average). In sales, however, total volume does not tell the entire story. Department stores expect to have some merchandise returned. The amount of percentage of sales returned is one measure of sales success. In this same department, for this same month, returned merchandise records were studied. As Fig. 1-5 indicates, one employee had only 4.5 per cent of her dollar sales returned, while one other had 21.3 per cent of her dollar sales returned. All other employees were between these extremes, and the average return was in the vicinity of 9 per cent. Regardless of whether sales volume or return percentage is considered, there are extensive differences between sales employees.

Characteristics of Production Data.—The examples given above, purposely drawn from unlike jobs, demonstrate three specific facts. (1) Production records (quality, quantity, or other kinds) point to real differences in job performance. (2) The magnitude of the difference between the best performance and the poorest performance is in the ratio of approximately

two to one *or higher,* depending in part on whether or not inexperienced employees have been considered. (3) A distribution of the performance of a group of employees follows a characteristic bell-shaped pattern with the largest group in the center, near the average, a small group of exceptionally good employees at one end, and another small group of exceptionally poor em-

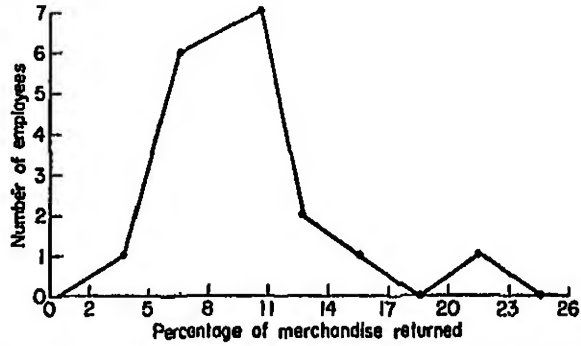


FIG. 1-5.—Distribution of percentage of merchandise returned for a group of curtain and drapery saleswomen in a department store.

ployees at the other, *even when beginners have not been considered.* These are well-known psychological facts that manifest themselves in job performance just as they do in the psychological laboratory and in all other areas of human behavior.

Determiners of Differences.—It is not within the province of this book to present an involved theoretical treatment of the factors that operate to produce these ability or performance differences that exist among people. Perhaps it is sufficient to point out that the multiplicity of hereditary and environmental influences interacting with each other result in these human differences. It is important, however, to note that any individual's performance may be improved through better training and better work methods, that his performance may be either improved or hindered by changes in lighting, hours of work, and other specific job changes. However, the differences that have been mentioned above cannot be explained away in terms of these factors. Under

given lighting conditions the bell-shaped distribution will be found. A change in lighting conditions *may result in increased production;* everyone may improve, but the characteristic pattern will remain.

Other Employee Differences.—Some employees work less than a week at a new job and then quit. Others work a little

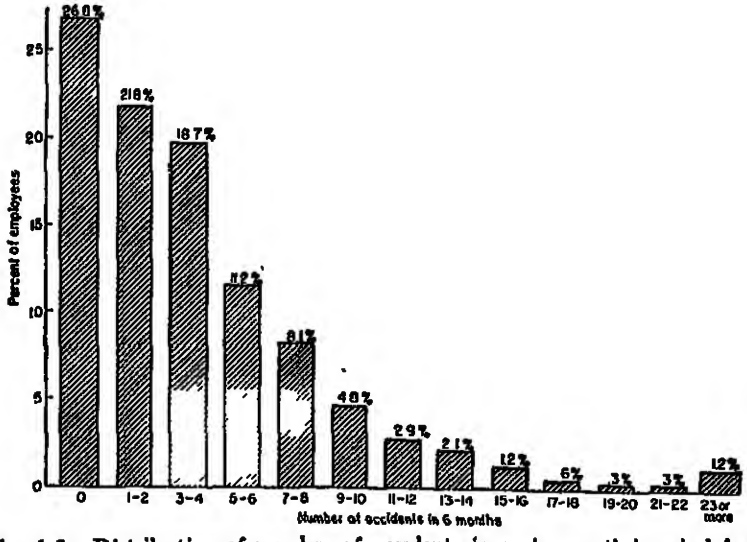


FIG. 1-6.—Distribution of number of accidents in a six months' period for 680 employees in an automobile-manufacturing plant. (*From Newbold. By permission of the Controller of His Britannic Majesty's Stationery Office.*)

longer and then decide to leave. Most company records indicate that the greatest likelihood of losing an employee occurs in the first few days or weeks of employment and that the longer an employee is on the job, other things being equal, the greater the probability of his staying with the company. A curve, then, constructed with work periods on the base line, designed to show the percentage who quit in the first period after employment, the percentage who quit in the second period after employment, and so on, would assume a shape characterized as being high on the left and tapering off toward the base line as it moved to the right. This kind of curve is characteristic of certain types of data including accident and absentee data.

Accident Records.—Figure 1-6 taken from an English study[1] is typical of most accident distributions. In an automobile-manufacturing plant the records of 680 employees for a six months' period were examined. As the figure shows, 26.8 per cent of the group had no accidents at all; 21.8 per cent had one or two; 18.7 per cent had three or four; and so on, the greater the number of accidents, the smaller the percentage of employees falling in the category.

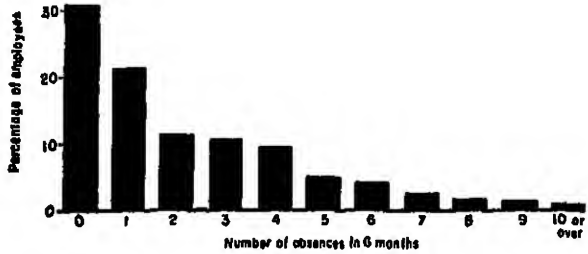


FIG. 1-7.—Distribution of number of absences in a six months' period for 151 employees in a casting shop. (*From Fox and Scott.*)

Absentee Records.—Absentee records show a similar distribution. Figure 1-7 shows the distribution of absences among 151 employees[2] in a casting shop for a six months' period. About 31 per cent of the group had no absences during the period studied; about 22 per cent had one; and so on, the more absences, the fewer the employees falling into that category.

What Is a Good Employee?—Every management must decide what it expects from its employees. Nearly everyone, however, would agree that those who are accident free, who come to work regularly, who stay on the job a reasonable length of time, who produce a lot of goods or sell a lot of merchandise, and who produce goods of high quality or sell merchandise that is not returned are the best employees. Which criterion is most important or how relatively important each is, is a question that

[1] NEWBOLD, E. M. *A contribution to the study of the human factor in the causation of accidents.* Great Britain Medical Research Council, Research Board, Reprint No. 38, 1926.

[2] FOX, JOHN B., AND SCOTT, JEROME F. *Absenteeism: management's problem.* Business Research Study No. 20, Vol. 30, No. 4, December 1943. Boston: Harvard University, Bureau of Business Research.

must be decided in the light of each company's own operations and its own policy.

Why Test?—One fact, however, is universal. To the extent that a company can select more and more employees who fall at the better end of these distributions and fewer and fewer who fall at the poorer ends of the curve, the better the result. In the case of the grocery checkers, for instance, if by means of personnel tests it were possible to eliminate, prior to hiring, all checkers who require more than 400 seconds (after six months) to check the standard order upon which Fig. 1-3 is based, the effect would be far-reaching. Either the store in question would need fewer checkers, or customers could be passed through the checking stands at a faster rate of speed, thus reducing time in line and improving customer good will. The best personnel on any job means lower costs, better customer relations, and fewer managerial and supervisory problems. *Personnel testing can contribute to these objectives.*

What Tests Will and Will Not Do.—The question of what tests will do and what they will not do can best be answered by this whole book. A few words in the nature of a preview are in order, however. Tests are not a cure-all for all personnel ills. They cannot be used to clean up the results of mismanagement and supervisory bungling. They will not always work in every situation; and when they do work, they will not yield perfect results. The adequacy of a test or a testing program is evaluated, not in terms of perfection but in terms of batting odds. A particular test should not be criticized because it resulted in the hiring of one or two bad employees but rather should be evaluated in terms of whether or not it selected *fewer* bad employees than the previously used technique.

Tests as an Aid.—And finally it should be clearly understood that tests are *not* advocated as a substitute for tried and true selection and placement procedures. Instead, tests are instruments that yield facts about the applicant, which facts, in combination with other facts obtained from the application blank, from references, and from the interview, make possible a more intelligent and reliable hiring decision.

CHAPTER II

PROCEDURES FOR CHOOSING TESTS

Since it is clearly apparent that employees do differ in the degree to which they measure up to the requirements of their jobs, the basic problem becomes one of selecting employment and placement tools that will aid in increasing the number of desirable employees and decreasing the number of less desirable employees.

Tests Sometimes Misrepresented.—In many instances, tests have been misrepresented. Individuals inexperienced in their use have been led to believe that they will accomplish far more than can rightfully be expected of them. First of all, *no single test or combination of tests will ever do a perfect job.* There are far too many personality factors contributing to successful job performance that are not measured by presently available tests. An individual may have the capacity or ability for performing a particular job; however, if there are elements in his home life that keep him in a constant state of emotional turmoil, he is not likely to measure up to the level predicted by the test. For this and hundreds of other reasons like it, tests do something less than a perfect job.

Figure 2-1 illustrates this fact effectively. A test battery[1] consisting of three tests was set up for the purpose of identifying potentially successful naval electrical trainees. The scores on the three tests were combined into a single value which, in turn, was used to predict success in the electrical training program as measured by school grades. The prediction was about as accurate as can be expected with present tests, and yet the results were

[1] LAWSHE, C. H., JR., AND THORNTON, G. R. A test battery for identifying potentially successful naval electrical trainees. *J. appl. Psychol.*, 1943, 27, 399–406.

something less than perfect as the figure shows. Each bar represents 20 per cent of the trainees, with the group having the highest combined test scores at the top and the group having the lowest scores at the bottom. The shaded portion of each bar represents the percentage of the group that received school grades below the average of the whole group, whereas the solid

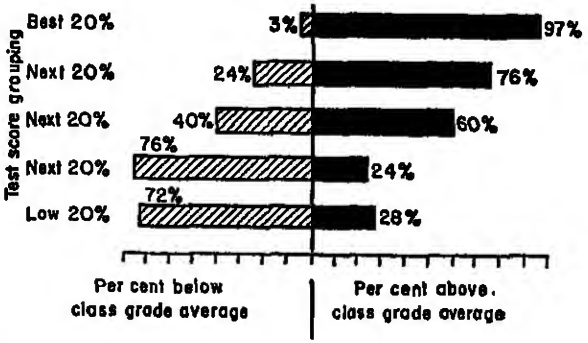


FIG. 2-1.—Naval electrical trainees assigned to five groups on basis of combined test scores to show proportions above and below average class grade.

section indicates the percentage with grades above average. In spite of the fact that the prediction here is as good as is usually found, among the best 20 per cent on the test, 3 per cent fell below average in school performance. Tests sometimes indicate the selection of an individual who does not turn out well on the job; sometimes applicants who would have become good employees are turned away. It must be borne in mind, however, that the same statements can also be made with respect to the interview or any other device that is used to select or allocate employees or applicants. The question becomes one, therefore, not of whether or not tests do a perfect job but of how much better than present methods they are. It must constantly be remembered, however, that almost *never* will a test or combination of tests predict so accurately that every applicant who scores above a given point is sure to succeed, while every applicant who scores below that point is sure to fail.

Standards Based on Opinion.—A second result of misrepresentation has come about when certain experts have professed to

be able to set up test batteries on the basis of subjective analyses alone. Such a procedure, unless it is accompanied by accepted fact-finding techniques, is an extremely hazardous approach regardless of whether it is employed by an outside expert or by an employee of the company. Frequently, the adequacy or inadequacy of the recommendations that are made is never known.

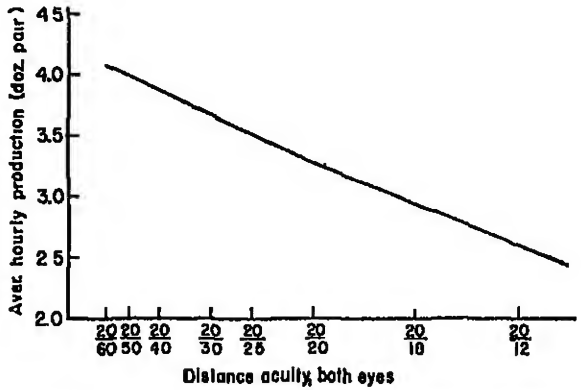


FIG. 2-2—Average number of dozens of pairs of hose processed per hour by employees performing at various levels on a distance acuity test. (*From Tiffin and Wirt.*)

How well these who were turned away might have done is a question that is rarely asked, much less answered. Sometimes these guesses result in the selection of persons who are actually poorest on the job. An example from the field of vision testing is a case in point.

Looping is an operation in the making of hosiery and involves meticulous activity on the part of the operator[1] at about eight inches from the eyes. The relationship between the production records of 199 loopers and their scores on the standard Snellen chart vision test at twenty feet is shown in Fig. 2-2. As shown here, those employees with the best distance acuity scores actually looped fewer dozens of pairs of hose than did those who

[1] TIFFIN, JOSEPH, AND WIRT, S. E. Near versus distance visual acuity in relation to success on close industrial jobs. *Supplement to Trans. Amer. Acad. Ophthal. and Otolar.* Rochester, Minn., June, 1944.

made a poorer showing on the Snellen chart. Discussion of this fact is postponed to Chap. VIII, but the example serves as an excellent illustration. To take the point of view that "since the job of looping requires 'good' eyes, we should select those applicants who can read the Snellen chart the best," actually resulted in the best applicants being turned away. Testing programs based entirely upon guesses or estimates sooner or later fall by the wayside. The best technician, the personnel manager, or the director of industrial relations eventually is asked questions that he cannot answer unless he is fortified with the facts. Sometimes these questions are raised by top management, sometimes by supervision, and sometimes by the union. Facts are just as essential to the operation of a personnel testing program as they are to the operation of a life-insurance company. Without a knowledge of mortality rates at different ages, for example, an insurance company could not operate. Without the facts regarding the probabilities of improvements resulting from test usage, the testing program is doomed to failure.

Two Techniques.—There are two basic fact-finding techniques whereby it is possible to know whether or not a given test or combination of tests should be used as a tool in allocating personnel to a particular job. Both of these are known as *validating* techniques and have as their purpose testing the test. A given test may be excellent in connection with one job and virtually useless in connection with another job. Furthermore, job classifications that seem similar from plant to plant sometimes differ significantly; so it becomes essential to test the test in practically every new situation. Such validation procedures simply answer the question "Does this test aid in identifying those persons who are most apt to be successful on this particular job?" These methods are referred to as the *present employee method* and the *follow-up method.*

PRESENT EMPLOYEE METHOD

The present employee method is most frequently used and gets results most quickly. It consists of five simple steps:

1. Analyze the job
2. Select a trial battery
3. Identify criterion groups
4. Administer the trial battery
5. Compare test results

Analyzing the Job.—The particular job in question should be studied for the purpose of estimating the specific demands that it seems to place upon the employee. Is the employee required to read blueprints; must he be able to perform the common arithmetic operations; does extreme finger dexterity seem to be required? The extent and formality of the job analysis will depend upon the test technician's familiarity with the job. The objective, however, is to list the apparent abilities, skills, and attributes that are measurable with present test methods.

Selecting the Trial Battery.—The next step is to select a number of tests that are intended to measure the traits or attributes reflected in the job analysis. The list of tests presented in Appendix C should be sufficiently inclusive for most purposes. However, the *Mental Measurements Yearbook*[1] may also be useful. The number of tests or the number of hours of testing to be included in the trial battery will depend upon the availability of the employees for testing as well as the relative importance of adequate placement on the job in question. The more important it is to have good employees on a particular job, the more time a company can afford to put into the development of an adequate test battery. In the case of bakery routemen, for example, poorly selected men not only result in lower sales but sometimes are the cause for losing a particular grocer as a customer, a condition that cannot easily be rectified. The general rule is that the more serious hiring mistakes can be, the more time the company can afford to spend in setting up the test battery. Generally speaking, however, three hours of well-chosen tests will give satisfactory results.

[1] Buros, Oscar K. (Ed.). *The 1940 mental measurements yearbook.* Highland Park, N.J.: Mental Measurements Yearbook, 1941. 674 pp.

Identifying Criterion Groups.—The next step consists of selecting two groups from the ranks of the employees who are presently performing the job, one consisting of employees who are considered satisfactory in the performance of the job and the other composed of employees who are falling short of the demands being made of them. Various measures of job success for classifying present employees are treated in Chap. III. The important thing is to select two criterion groups, one consisting of those considered satisfactory and the other consisting of those considered generally unsatisfactory.

Administering the Trial Battery.—Once these two groups of employees have been identified, they should be called together in the conference room, the cafeteria, or some other suitable meeting place (see Appendix B) and asked to take the tests. They should *not* be told that some of them have been selected because they are performing the job satisfactorily and that others have been chosen because they are not doing so well. They should be advised, however, that the company in its efforts to find ways of placing the right people on the right jobs in the future is presently engaged in testing the tests. Key points well worth keeping in mind are:

1. No employee should be compelled to participate.
2. Each employee should be given the opportunity to withdraw without embarrassment if he so desires.
3. It should be made clear that no one's status with the company will be affected by his performance on the tests.

In the event that the employees in the company are organized, certain precautions should be taken. Although it is management's prerogative to decide whether or not it will use tests as a selection and placement tool, management does need the cooperation of the union when it is in the process of validating its tests by the present employee method. For this reason, it is well to contact the shop committee ahead of time. Such a contact should not be made in the spirit of "will you permit us to do this?" but rather in the spirit of informing the employees. *Managements that disregard this suggestion frequently fail in their testing pro-*

grams. Without the proper information in the possession of the employees, trouble usually results. Where labor relations have previously been good and where the story is honestly and adequately presented, cooperation usually follows. In those rare instances in which this is not the case and the union chooses not to cooperate, there is usually nothing to be gained by forcing the issue. In these instances it usually pays to substitute the follow-up method discussed later.

Comparing Test Results.—Various methods for analyzing and presenting the results of this tryout procedure are discussed in Chap. IV. However, the most simple and obvious approach will be discussed here. Suppose that having selected and tested two criterion groups, it appears that both groups average about the same on a given test. It can be concluded that this test has no value for purposes of identifying the potentially successful employees *on this job.* By this very simple procedure or one of the variations of it discussed in Chap. IV the facts can be collected. *These facts are absolutely essential if a program is to succeed.*

THE FOLLOW-UP METHOD

In reality, the follow-up method is only a modification of the present employee method. As the name suggests it involves the use of newly hired employees *not* presently on the job. It likewise consists of five steps, only two of which differ from those in the above method.

1. Analyze the job
2. Select a trial battery
3. Test all new hires
4. Classify them into criterion groups
5. Compare test results

Testing New Hires.—Since the first two steps are identical in the two methods, no further discussion of them seems necessary. This procedure calls for the administration of the trial battery to all new people who are placed on the job. This testing can be integrated with the usual hiring procedure, and no comment need be made to the employee regarding the use of tests for the first

time. It is well, however, to follow the suggestions presented in Appendix B so as to ensure maximum cooperation from the applicant. Most important of all is the fact that *during the tryout period test scores should not be known by interviewers* or others involved in the hiring procedure. Extreme care should be exercised to prevent the test scores from influencing those responsible for hiring. This is because it is not known at this time whether or not the test is correlated with job success. At no time should the person responsible for testing lose sight of the fact that this is a tryout period and that scores should *not* be used until the facts are known. Tests and scores should be filed for use as indicated below.

Classifying the New Employees.—After a sufficient number of new employees have been tested and placed on the job and after sufficient time on the job has elapsed to permit judgments to be made or records to have accumulated, these employees should be segregated into criterion groups in the same fashion as suggested in connection with the present employee method. One of the approaches outlined in Chap. III will usually suffice. The question of when an employee has been on the job long enough is sometimes a perplexing one. The time period will vary with the type of operation and the nature of the training program. In operations where individual production data are available it is sometimes wise to plot the average performance by days or weeks during the initial periods of employment. Figure 2-3 shows such a curve for five employees learning to inspect ophthalmic lenses. Even though these employees began at different times, their records were grouped for their first day on the job, their second day, etc., and hourly averages were computed. Although this particular learning curve does not cover a sufficient time period to permit a valid judgment to be made as to whether or not the initial learning period has passed, nevertheless it is typical of such learning curves in that it is characterized by a rapid initial rise plus a gradual flattening off. On this particular job, learning seems to be reasonably rapid. However, on many jobs where learning is slower the base line can be more advantageously plotted by weeks or even months. The point is that *no attempt to categorize new*

employees into criterion groups should be made until there is some evidence that the initial learning period has passed.

Comparing Test Results.—Once these new hires are classified into criterion groups, the procedure is then identical with that used in the present employee method. The scores on a given test are averaged for each criterion group, and the averages of the two groups are compared. Only if there are significant differences in

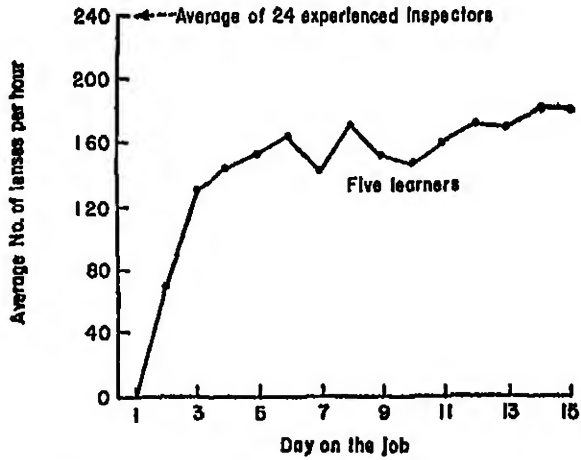


FIG. 2-3.—Average number of ophthalmic lens inspected per hour by learners during their first few days on the job.

the average scores of the two groups does the test differentiate between good and poor employees. Attention is called to other methods for analyzing and comparing data presented in Chap. IV.

COMPARISON OF METHODS

There has been no intent to imply here that the use of either method of test validation presupposes the exclusion of the other method. Probably the ideal approach is to use both methods, checking the results, one against the other. Each method, however, has its advantages and disadvantages.

Advantages of Present Employee Method.—Perhaps the greatest advantage of the present employee method is the speed with which results can be obtained. If hiring is slow, it may take six or eight months before a sufficient number of new people have

been hired to permit a comparison by the follow-up method. Results may be obtained in a matter of hours with the present employee method.

Advantages of Follow-up Method.—(1) The primary advantage of the follow-up method lies in the fact that the amount of selling which needs to be done prior to inaugurating the program is greatly minimized. It is not nearly so necessary to get prior support from supervision, the union, or others. All of that will come after facts have been accumulated, in which case the facts speak for themselves. (2) The testing situation in which applicants are tested during the tryout period is essentially the same as that in which subsequent applicants will be tested. There can be no objections based upon the real or imaginary argument that applicants are different from employed personnel. Occasionally there are instances in which the present employee method cannot be used. If the operation is a new one and there are no present employees or if men have always been employed in the past and women are to be hired now, obviously the follow-up method must be used.

Which to Choose.—The question of which method to use is one that must be answered in the light of the circumstances. If these circumstances permit a free choice of either, it is probably best to use both and to check the results, one against the other. Should the results of such a dual tryout differ, those resulting from the follow-up method should usually be accepted.

SUMMARY

Tests can contribute materially to better personnel selection and placement. The choosing of tests for a given purpose, however, must be done systematically and scientifically. Without actual facts indicating how well a test or group of tests identifies the superior employees, no intelligent judgment can be made as to whether or not that particular test should be used. Two general techniques for testing the test are available: the present employee method and the follow-up method. Each has its advantages, and the decision as to which should be used is one that must be made in the light of conditions existing in the particular plant or company.

CHAPTER III

MEASURES OF JOB SUCCESS

Identifying two groups of employees, one of which is known to be good on the job and another which is known to be poor on the job, is in many respects the most important step in the test-validation procedure. This is true regardless of whether the present employee method or the follow-up method is used. How well or how poorly this step is done will often determine how successful the chosen tests will be in the future.

Criteria of Success Vary.—The question of who is most successful on the job is associated with the nature of the operation. For example, an operator charged with the responsibility of finishing an aircraft propeller blade that has already had a thousand dollars worth of labor put into it makes a very costly error if he damages the blade so that it cannot be salvaged. How many he finished in a given period of time is relatively unimportant, the best measure of success on his job being the quality of his work as evaluated by the amount of damaged material. On the other hand, the best criterion of success in the assembly of a certain type of electrical fixtures would very likely be quantity. In a competitive market on products involving a relatively high labor cost it is perhaps more important to select operators who produce a lot, even at the expense of errors. Here the measure of the successful operator is in terms of how many he produces or assembles. On some jobs the most important item is the reduction of personal accidents; so the best employee is the accident-free individual, and the poorest employee is the accident-prone person. The terms "good" and "poor" applied to employees are relative and are matters of definition which must be made in the light of the critical factors inherent in the job itself.

Success a Matter of Policy.—Frequently this definition of success is a policy matter, independent from the operation itself. The definition of a good salesman, for example, varies considerably from company to company. Is he the person who makes the most calls, the one who secures the most new accounts, the one with the greatest volume, the one with the greatest volume on certain "long profit" items, or is he the one with the fewest customer complaints?

For example, one life-insurance company might build its business strictly on annual volume with little regard for cancellations as long as the volume is high. The management of another company might feel that volume alone does not tell the whole story. Such a company might prefer to sell less to each individual, keeping the amount within his ability to carry it, and make its quotas through sales to more individuals. It is clear that volume alone does not tell the complete story. "What is a good insurance salesman" can very well differ from company to company.

Availability of Facts.—Not the smallest consideration in selecting a criterion is the availability of facts. Entirely too often for comfort, the personnel test technician finds either that individual records on employees have not been kept or, if they have been kept, that they are not filed in a way which makes them available with a reasonable amount of clerical work. One service that can nearly always be performed by the personnel man is the encouragement of management and supervision to keep adequate records. Who is producing the most and who is making the mistakes is a question that can be answered all too infrequently.

Four Classes of Criteria.—Little can be said here regarding what standards or policies should be considered in setting up the two groups of employees for a specific job. The purpose of this chapter is to present a reasonably comprehensive list from which can be drawn measures that are most appropriate for a specific job or operation in a specific company. In general, there are four broad areas: (1) production data, (2) personnel data, (3) judgments of others, and (4) job samples.

PRODUCTION DATA

Production data, as the name implies, are facts of a quantitative, numerical sort that are already available or that can be made available in the future. As a measure of job success they are useful only where people work as individuals rather than in teams or on lines.

Quantity.—How many pieces or parts per hour, day, or week?

Time.—How long to do a particular job (usually applicable to service types of operations)?

Quality.—How many rejects or reworks? How many dollars worth of scrap? How many errors?

Earnings.—How many cents per hour (where piece rates or bonuses are in effect) did the employee earn? Did he earn a bonus?

These are perhaps the obvious ones. Others will suggest themselves as specific jobs are studied.

Sales Criteria.—Although the general problem of validating tests for salesmen is essentially the same as any other validation problem, the matter of sales criteria is by nature unique. Ohmann [1] has presented a comprehensive list from which the following ones have been adapted:

Sales volume for a given period of time
Average number of calls per day
Net commission earnings
Average number of sales per month
Average size of order
Average number of new accounts per month
Average sales volume per year for period of time employed
Sales volume for first six months on the job
Trend of sales volume over period of years
Amount of allowances to customers
Amount of returned merchandise
Classes of trade called on
Classes of products sold

[1] OHMANN, O. A. A report of research on the selection of salesmen at the Tremco Manufacturing Company. *J. appl. Psychol.*, 1941, **25**, 18–20.

Ohmann contends that in his own work, net commission earnings have been the best criterion, although there is no implication that any single criterion should be used at the exclusion of the others.

PERSONNEL DATA

Nonproduction Measures.—Separate and somewhat apart from how much or how well an employee produces are certain facts about him that deserve consideration in the identification of superior employees. For example, whether or not an employee misses work often is a factor to be considered, even though he is a good producer while he is on the job. Other things being equal, the employee who is always there is a better employee. These nonproduction facts usually involve the use of personnel data, some examples of which follow.

Absenteeism.—Who are the best attenders, and who are the poorest attenders? Fox and Scott [1] have reported that the number of absences per unit of time is a better measure than number of days absence per period of time, any continuous period off the job being considered an absence whether it is one day or two weeks.

Length of Service.—This is useful only when the follow-up method is used when employees can be classified into those who either terminated or were terminated within a six months' period, for example, and those who remained on the job six months or longer.

Rate of Advancement.—How long was the individual on the job before he was promoted to a better one?

Training Time.—How long did it take the individual to learn the job? Sometimes employees can be grouped into those who attained a given production level in a certain period of time and those who did not. (Sometimes this is classified under production data.)

Accidents.—Employees may be divided into those who had one or more accidents during a given period of time and those who had

[1] Fox, John B., and Scott, Jerome F. *Absenteeism: management's problem.* Business Research Study No. 29, Boston: Harvard University, School of Business Administration, 1943.

none during the same period. Lost time, home cases, hospital visits, or other indexes of severity may be utilized.

All of these measures likewise employ quantitative facts that are presumably available and are indicative of job success. To this list may be added others that are more or less specific for certain jobs; for example, in most service types of jobs, customer complaints are important. The list supplied above is general for most industrial jobs and will serve to suggest others.

JUDGMENT OF OTHERS

Frequency of Use.—The individual engaged in test validation will use supervisory judgments in the form of ratings, as a general rule, more frequently than he will use all of the other suggestions in this chapter combined. This is true because more times than not the job will be of such a character that individual records cannot be kept or, in those instances where records can be kept, they will not be available. Therefore, it is important to consider some of the common pitfalls of employee-rating methods and to examine means of eliminating or at least minimizing weaknesses that tend to reduce the reliability and the validity of the ratings.

Weaknesses of Scales.—Most merit-rating systems presently in use in industry, though better than nothing, are not sufficiently valid or reliable for use as criteria in testing the test. Some of the common weaknesses follow:

1. They do not take into account or correct for the individual differences among raters.
2. When several items are included, the "halo" effect is so great that the scales actually yield much less than they appear to.
3. No effort is made to compensate for job differences.

Differences between Raters.—The English language is not sufficiently exact for a given employee description to mean the same to everyone who reads it. In addition, the rater's own temperament is reflected in any mental standards that he may set. And so when two or more raters are asked whether an employee's work habits are poor, average, or excellent, different answers are obtained, even when the raters have had an equal opportunity to

observe the man's work. "Excellent" means different things to different people, and supervisors are no exception.

Figures 3-1 to 3-4 are hypothetical examples to illustrate dif-

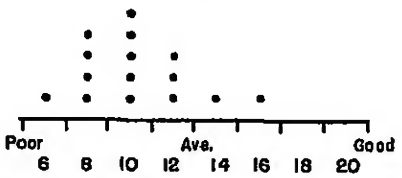


Fig. 3-1.—Distribution of ratings of fifteen men by Supervisor A, averaging 10.

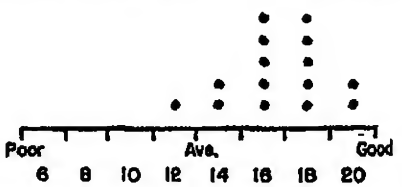


Fig. 3-2.—Distribution of ratings of fifteen men by Supervisor B, averaging 17.

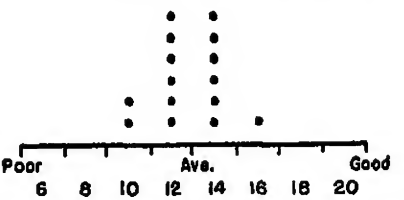


Fig. 3-3.—Distribution of ratings of fifteen men by Supervisor C, averaging 13.

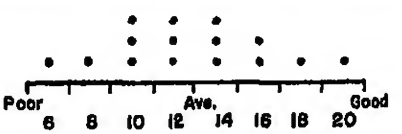


Fig. 3-4.—Distribution of ratings of fifteen men by Supervisor D, averaging 13.

ferent types of raters rating the same fifteen men. Rating values on this particular scale item range from 6 to 20 in two-point steps.

Note that Supervisor A (Fig. 3-1), who might be thought of as the "hard-boiled" variety, rated most of the fifteen at 8, 10, or 12 and that his ratings averaged 10.

Supervisor B, the "easy-going" variety (Fig. 3-2), rated most of his at 16 or 18 with an average for the fifteen ratings of 17.

Supervisor C (Fig. 3-3), the "Caspar Milquetoast" variety, being afraid to make decisions, rated all but three men at 12 and 14 and had an over-all average of 13.

Supervisor D (Fig. 3-4), using the scale to the best advantage, spread the men along the entire range of the scale but also had an average of 13.

Table I gives a summary of the four distributions and shows at a glance the fact that this scale can produce vastly different results in the hands of various supervisors. A rating of 10, for

TABLE I. — SUMMARY OF RATINGS OF FOUR SUPERVISORS FOR FIFTEEN MEN

Supervisor	Ratings		
	Highest	Average	Lowest
A	10	10	6
B	20	17	12
C	16	13	10
D	20	13	6

example, means average when given by A and very poor when given by C. A rating of 16 by A or C means very good but means below average for B. It can readily be seen that any pooling of ratings by different judges can mean chaos unless some approach is utilized to ensure greater uniformity. The training of supervisors will help some, but these differences which are a function of temperament as much as interpretation will always remain when not controlled.

While these examples are hypothetical for purposes of illustrations, they are not exaggerations and are truly indicative of what is found in industry. Figure 3-5 shows the actual distribution of ratings for three different departments in a steel mill. An eleven-item merit-rating scale was used; points were awarded on each item; and total points were converted to A, B, C, or D ratings for each employee. Note the discrepancy in the proportion of ratings from department to department. Note that an employee in the annealing and pickling department has eight chances in a hundred of receiving a D while no one in the other two received

D's. His chances of receiving either A or B are 78 out of a 100 in the transportation department and 45 out of a 100 in the annealing and pickling department. Although the ability of employees to do their jobs can vary from department to department, it is not reasonable that eight out of a hundred in one department are bad and that there are no bad employees in another. Ratings fre-

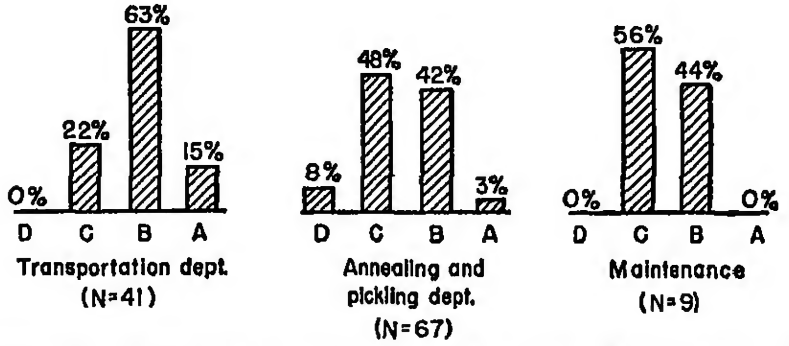


FIG. 3-5.—Distributions of merit-rating scores in three different departments of a steel mill showing the percentage of men receiving A, B, C, and D ratings.

quently tell more about the person doing the ratings than about the employees being rated.

Halo Effect.—When persons who prepare merit-rating systems include long lists of items such as initiative, cooperation, ingenuity, and so on, the implication is that the person doing the rating can look at an employee one way and see only "initiative," that he can look at him another way and see only "cooperation," and so on. Such is contrary to the fact as is demonstrated by research studies.[1] In reality, the supervisor thinks of an individual largely in terms of how well he believes he is measuring up to what is expected of him. In other words, job performance is what the supervisor sees, and he sees it when he is looking for initiative, when he is looking for cooperation, and when he is looking for ingenuity. It is not strange, then, that when a supervisor attempts to rate his employees on these complicated rating scales,

[1] EWART, EDWIN, SEASHORE, S. E., AND TIFFIN, JOSEPH. A factor analysis of an industrial merit rating scale. *J. appl. Psychol.*, 1941, 25, 481–486.

he tends to place a given employee at about the same position on each scale, regardless of what that particular scale is called or what it was intended to measure. Figure 3-6 shows the actual ratings of fifteen men drawn at random from some 9,000 employees in another steel mill. The rating scale [1] used employed eleven items, and the figure shows the relationship between the

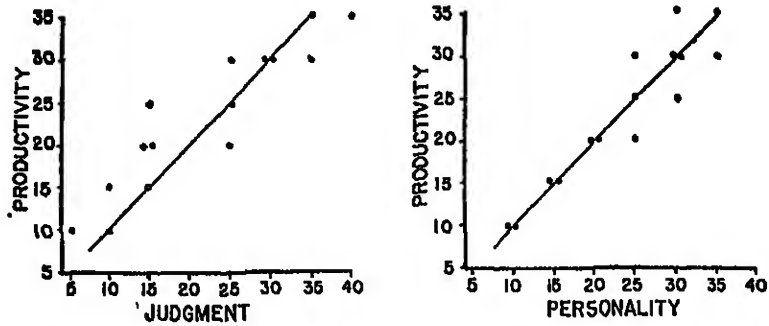


FIG. 3-6.—Scattergrams showing relationship between productivity and judgment ratings and productivity and personality ratings for fifteen men in a steel mill.

"productivity" item and the "judgment" and "personality" items. Note the reasonably close agreement. An individual who was rated at a given point on the productivity scale tended to be rated at about the same point on each of the other two. This is known as the "halo" effect and suggests that the many different items are, in reality, measuring about the same thing and that some greatly simplified system might do as well or even better.

Job Differences.—In theory, at least, it should be just as nearly possible to have a janitor who is doing a 100 per cent job as it is to have a tool- and diemaker who is doing a 100 per cent job. Figure 3-7 shows five jobs selected at random from Tiffin's longer list.[2] Actually, the average number of merit-rating points in each job classification decreases as the skill level of the job decreases. This suggests that these employees have not really been

[1] TIFFIN, JOSEPH, AND MUSSER, WAYNE. Weighting merit rating items. *J. appl. Psychol.*, 1942, 26, 575-583.

[2] TIFFIN, JOSEPH. *Industrial Psychology*. New York: Prentice-Hall, 1947, p. 354.

measured against the demands of their respective jobs but that their general competence as individuals has been evaluated without reference to job classification.

Suggested Plans.—Best results have been obtained by using one or both of two very simple plans, either of which will overcome most of the weaknesses cited in connection with usual merit-

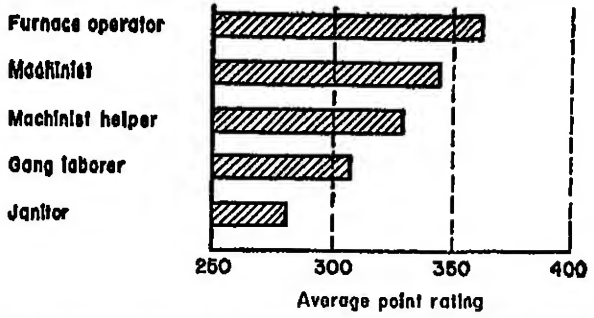


FIG. 3-7.—Average total point ratings for men in five job classifications in a steel mill. (*From Tiffin.*)

rating plans. One is called the *card-stacking method*, and the other is called the *ranking method*. Fundamentally they are the same and differ only in certain minor respects.

The Card-stacking Method.—After the supervisor who is to do the rating has been led to see the need and value of rating people on the job for purposes of test validation, he is supplied with a stack of cards similar to the one shown in Fig. 3-8, each of which bears the name of one employee under his supervision. Next, he is asked to make three rough groupings of these cards, placing in one stack his superior employees, in one his below-average employees, and in the other his average people. If he argues that he has no above-average employees, for example, he is reminded that "some employees are better than others" and that he should put the best he has in one stack. After this is done, reference is made to Table II and he is asked to make a second grouping. For example, if he has placed more than 30 per cent in the best group, he is asked to identify the poorest so-many and to move them to the middle group. In other words, his first grouping is corrected so that he has 30, 40, and 30 per cent in the re-

TABLE II.—FOR DETERMINING NUMBER OF EMPLOYEES TO BE PLACED IN EACH CATEGORY WHEN USING THE CARD-STACKING METHOD FOR RATING

Second grouping				Third grouping					
No. rated	Best 30%	Middle 40%	Poorest 30%	No. rated	Best 10%	Next 20%	Middle 40%	Next 20%	Poorest 10%
10	3	4	3	10	1	2	4	2	1
11	3	5	3	11	1	2	5	2	1
12	4	4	4	12	1	3	4	3	1
13	4	5	4	13	1	3	5	3	1
14	4	6	4	14	1	3	6	3	1
15	5	5	5	15	2	3	5	3	2
16	5	6	5	16	2	3	6	3	2
17	5	7	5	17	2	3	7	3	2
18	6	6	6	18	2	4	6	4	2
19	6	7	6	19	2	4	7	4	2
20	6	8	6	20	2	4	8	4	2
21	6	9	6	21	2	4	9	4	2
22	6	10	6	22	2	4	10	4	2
23	7	9	7	23	2	5	9	5	2
24	7	10	7	24	2	5	10	5	2
25	7	11	7	25	2	5	11	5	2
26	8	10	8	26	3	5	10	5	3
27	8	11	8	27	3	5	11	5	3
28	9	10	9	28	3	6	10	6	3
29	9	11	9	29	3	6	11	6	3
30	9	12	9	30	3	6	12	6	3
31	9	13	9	31	3	6	13	6	3
32	9	14	9	32	3	6	14	6	3
33	10	13	10	33	3	7	13	7	3
34	10	14	10	34	3	7	14	7	3
35	11	13	11	35	4	7	13	7	4
36	11	14	11	36	4	7	14	7	4
37	11	15	11	37	4	7	15	7	4
38	12	14	12	38	4	8	14	8	4
39	12	15	12	39	4	8	15	8	5

spective stacks. Finally, again using Table II, he is asked to make a third grouping so that he ultimately has five stacks of cards containing 10, 20, 40, 20, and 10 per cent of the total group, respectively. He then makes an x in the appropriate space at the

HOW TO OBTAIN SUPERVISORY RATINGS FOR USE AS TEST CRITERIA

STEP 1.—*Get supervisor to see need for rating*
Use group conference or individual interview
Explain what he has to gain
Keep approach informal
Explain confidential nature
Emphasize *present job*

STEP 2.—*Provide him with typed cards*
Three by five inches
One name to a card

STEP 3.—*Have him make three rough groupings*
"Some employees are better than others"
One pile "best"
One pile "poorest"
One pile "in between" or "average"
Throw out any "don't knows"

STEP 4.—*Correct distribution to* 30, 40, 30 *per cent* (second grouping)
Use informal approach
"Who is poorest in this group?"
"Who is best in this group?"

STEP 5.—*Have him extend the distribution to* 10, 20, 40, 20, 10 *per cent* (third grouping)
Have him identify the extreme employees
"Who are the best *so many* in the high group?"
"Who are the poorest *so many* in the low group?"

STEP 6.—*Record ratings on cards*
Use A, B, C, D, E or
Use 5, 4, 3, 2, 1

bottom of the card and dates and initials or signs it. Letter ratings of A, B, C, D, and E or numerical ratings from 1 to 5 can later be assigned. The system works well when the supervisor has ten or more employees. A simplified step-by-step procedure is presented on page 31.

The Ranking Method.—The ranking method is quite similar except that it is better adapted to small groups of fifteen or less.

Name _____ _____ Dept. _____
Last First

Job Performance

How satisfactory is this employee in the performance of his present job?

NOTE- Consider such items as general productivity, including quality and quantity of work, specific job knowledge, safety, industry, dependability, cooperation, and perseverance.

Poorest 10%	Next 20%	Middle 40%	Next 20%	Best 10%
☐	☐	☐	☐	☐

Date _____ Rated by _____

FIG. 3-8.—Sample form for use in rating employees by the *card-stacking method.*

The supervisor is likewise handed a pack of cards, each one bearing an employee's name (see Fig. 3-9). In this instance, since the number is small, he is simply asked to rank the cards, placing the one bearing the best employee's name on the top. Since extremes are more easily identified, it is best to ask him to select the best and poorest first. Then, among those remaining, select the best and the poorest and so on. Entries are made on the card to indicate the number rated and the employee's rank number. Later, by means of Table III, a value from 1 to 5 may be assigned to make the ratings roughly comparable to those obtained by the card-stacking system.

Forcing the Distribution.—Both of these systems are intended to "force the distribution," that is, to force the rater to use the

Name _____ _____ Dept _____
Last First

Job Performance

How satisfactory is this employee in the performance of his present job?

NOTE Consider such items as general productivity, including quality and quantity of work, specific job knowledge, safety, industry, dependability, cooperation, perseverance

Number of employees ranked — — — — — — — — — — — — (_____)

This employee's rank (No 1 is high) — — — — — — — — — — (_____)

Leave blank — — — — — — — — — — (_____)

Date _____ Ranked by _____

FIG. 3-9.—Sample form for use in rating employees by the *ranking method.*

TABLE III.—FOR CONVERTING EMPLOYEE RANKS TO A FIVE-POINT SCALE

Rank position	Number ranked											Rank position
	5	6	7	8	9	10	11	12	13	14	15	
1	5	5	5	5	5	5	5	5	5	5	5	1
2	4	4	4	4	4	4	4	4	4	4	5	2
3	3	3	3	4	4	4	4	4	4	4	4	3
4	2	3	3	3	3	3	3	4	4	4	4	4
5	1	2	3	3	3	3	3	3	3	3	4	5
6		1	2	2	3	3	3	3	3	3	3	6
7			1	2	2	3	3	3	3	3	3	7
8				1	2	2	3	3	3	3	3	8
9					1	2	2	2	3	3	3	9
10						1	2	2	2	3	3	10
11							1	2	2	2	2	11
12								1	2	2	2	12
13									1	2	2	13
14										1	1	14
15											1	15

whole range of the scale instead of some particular part. Without having to interpret such words as "excellent" and "poor," the supervisor identifies the best 10 per cent now in his department,

the poorest 10 per cent, etc. The spreading of the distribution is important because if all employees are rated the same or nearly the same, there is no differentiation, and the ratings are useless for test-validation purposes.

Pooling Judgments.—The general rule is that pooled judgments are superior to individual judgments, that the average rating of two judges is superior to the rating of either alone. This is true, however, *only when both judges are equally qualified to evaluate the employee.* So often in industry to secure an additional judge it is necessary to move to the next higher level of supervision, in which case it is almost invariably true that the additional judge knows less about the employee's day-to-day competence. There are instances, however, where several supervisors are equally qualified to judge. For example, if employees rotate from shift to shift and supervisors do not, each supervisor has presumably equal opportunity to observe. Considerable judgment must be exercised by the personnel man in deciding what ratings to use. It is never wise to overlook completely the general competence of the supervisor in deciding whether or not to use his ratings.

JOB SAMPLES

Need for Job Samples.—Occasionally there are job classifications for which neither production nor personnel data are available and for which judgments by supervisors or others are not dependable. Many inspection types of jobs fall in this particular category. Since, as a general rule, the inspector is the last employee to handle the product, it goes on directly to the customer. If the inspector makes a mistake, the only way in which it is known is through a customer complaint. Very often the complaint cannot be traced to the person passing the defective or faulty item. In this kind of situation, supervisors have little or nothing to go on. They do not possess information that they can use in intelligent rating. Tiffin and Rogers' assorting room study [1]

[1] TIFFIN, JOSEPH, AND ROGERS, H. B. The selection and training of inspectors. *Personnel*, 1941, 18, 14–31.

is a case in point. In seeking a criterion against which to validate test scores they secured supervisory ratings. These ratings showed virtually no relationship with the tests used. Later, the operators who inspected tin plate for defects and sorted it were each asked to sort a standard stack, each sheet of which had been previously coded as "satisfactory" or "defective." Each girl was timed, and each was given a percentage accuracy score. The average time and accuracy scores for the twelve girls rated best by the supervisors were compared with those of thirty-eight girls

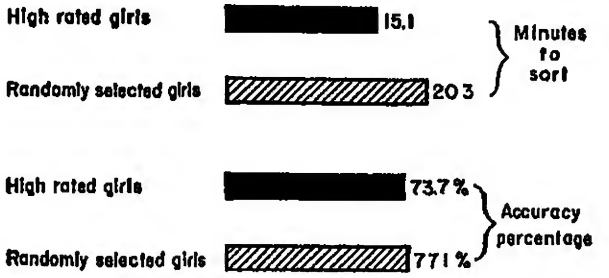


FIG. 3-10.—Performance of high rated sorters and randomly selected sorters on tin-plate job samples.

selected at random, and the results are presented in Fig. 3-10. Although the girls rated superior were slightly faster, requiring an average of 15.1 minutes as compared with 20.3 for the randomly selected girls, their accuracy was slightly lower. In other words, in the absence of other facts upon which to base judgment, the supervisors had evidently selected as superior those girls who handled the most metal in a day's time.

Nature of Job Sample.—A job sample is a portion of the job standardized in such a fashion that everyone performs the identical task. In the case cited above the same 150 sheets of tin plate sorted by each operator constitute a job sample. Another job sample reported by Lawshe and Tiffin [1] involved the performance of 200 inspectors with those instruments in a list of twenty that they were called upon to use in their jobs. Figure 3-11 shows the

[1] LAWSHE, C. H., JR., AND TIFFIN, JOSEPH. The accuracy of precision instrument inspection. *J. appl. Psychol.*, 1945, 29, 413–419.

percentage meeting the standard on eleven of the twenty job samples. Here objective measures of an employee's ability to do

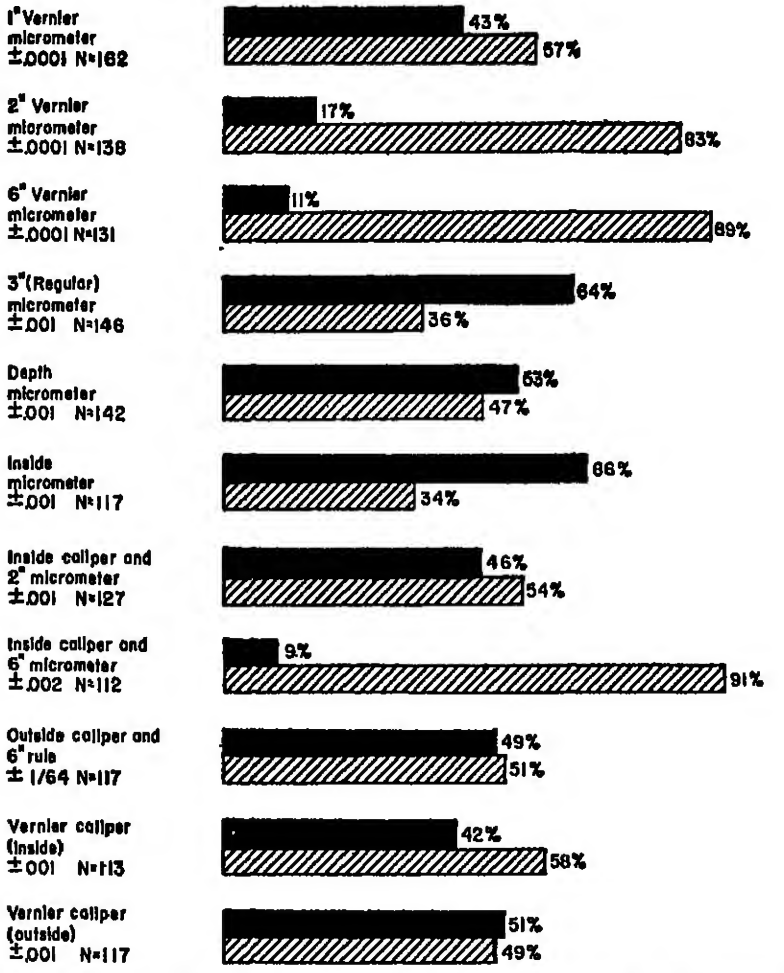


FIG. 8-11.—The percentage of inspectors passing and failing various precision-measuring instrument performance tests in an aircraft propeller plant. The solid bars indicate the percentage passing. (*From Lawshe and Tiffin.*)

his job or at least one aspect of it, while obtained for another purpose, could easily be utilized for segregating these employees into two groups for purposes of test validation. Whenever a por-

tion of the job can be pulled out and set up as a standardized task that everyone can be asked to perform, it can be thought of as a job sample.

THE SINGLE VARIABLE

Learning Time.—On page 17 in Chap. II, reference was made to the need for controlling the experience or learning factor in selecting criterion groups. If some of the employees have not been on the job long enough to learn it, they will probably have

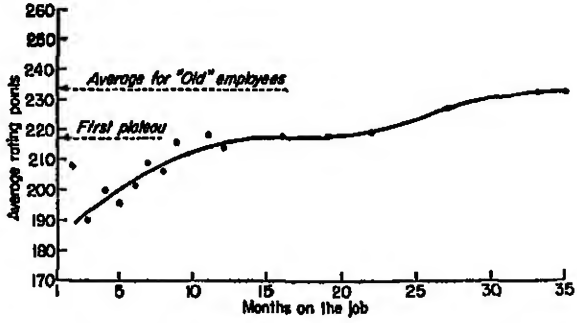


FIG. 8-12.—Curve showing trend of average merit-rating points for 1,000 steel mill employees by amount of service on the job.

low production or be rated low by supervisors. This, of course, places them in the poor group, when in reality there is no way of determining at that time whether they will actually be poor or good once they have had the opportunity to learn the job. As pointed out earlier, it is sometimes desirable to construct learning curves similar to Fig. 2-3 in order to get an idea of the length of the learning period.

The same phenomenon sometimes appears in connection with merit ratings. Figure 3-12 shows the average merit-rating points for one thousand employees in a steel mill by months with the company. It is quite evident that the tendency is for new employees to be rated lower than old employees. Whether the older employees are really better or supervisors only feel that they are better cannot be decided from these facts. However, the point is that in setting up criterion groups based upon ratings, it will quite likely happen that the good group would tend to have a

preponderance of older employees, whereas the newer group would tend to have a preponderance of newer employees. If this is the case, it is usually desirable to eliminate from the study those employees having extreme amounts of experience. In the case cited, one would be reasonably safe in eliminating all employees with fewer than twelve months' experience or more than twenty-four.

TABLE IV.—MEAN TEST SCORES OF FREIGHT SOLICITORS DIVIDED INTO RATING GROUPS

Rating	All solicitors		After elimination	
	N	Mean	*N*	Mean
A	21	−158	15	−158
B	22	−148	17	−134
C	12	−153	5	−125

If he did this, he could be certain that such differences as he might find in test scores between good and poor employees were not associated in any way with experience on the job. Likewise if he found no differences, he could be certain that true differences were not being masked by the experience factor.

Age and Other Factors.—In addition to experience, such factors as age, sex, and color should be controlled by eliminating a sufficient number of cases to make the criterion groups comparable. A case involving railway freight solicitors shows how a number of variables can mask the true facts. Forty freight solicitors who had been rated as A, B, and C were given the Bernreuter's *Personality Inventory* and their mean scores on the "neurotic tendency" (Bl-N) component are shown on the left side of Table IV. Systematic age and experience factors were apparent, however; and when all solicitors with less than two year's experience and all who were over fifty years of age were eliminated, thirty-seven remained, and their mean scores are shown on the right side of Table IV. Note that although no important differences are apparent when all are considered, there is a systematic trend to the averages when the select group is used.

SUMMARY

Selecting groups of employees that are relatively good and poor is one of the most important steps in the test-validation procedure. Generally speaking, the nature of the criterion of job success varies with the nature of the operation, the policies of the management, and the availability of facts. The four types of criterion data are production data, personnel data, judgment of others, and job sample performance.

CHAPTER IV

METHODS FOR ANALYZING AND PRESENTING FACTS

Throughout Chap. II the importance of "testing the test" is emphasized, and in the discussion of the five-step procedure reference was made to the comparison of the test results of the two criterion groups of employees. There are four basic approaches to the study of relationships between test scores and measures of job success: the scattergram, the method of averages, the method of percentages, and the profile method. Each of these has variations that may be adapted to the needs of particular problems. In order to illustrate the application of these several techniques, data collected in conjunction with the trade training program[1] for electricians mentioned earlier were used. Two hundred trainees were administered nine different tests prior to their admission to a fifteen weeks' training program. At the close of the program, each received an achievement grade somewhere between 2.0 and 4.0, the latter being the maximum that could be obtained. Test scores and grades in the training program are used here to illustrate the major methods of analysis and interpretation.

THE SCATTERGRAM

Two Variables.—Any time that two variables (that is, test scores and some measure of job success) are being studied, some estimate of their relationship can be made by preparing a graphic picture of this relationship, known as a scattergram. As is shown in Fig. 4-1, the vertical axis (called the *ordinate*) is used to

[1] LAWSHE, C. H., JR., AND THORNTON, G. R. A test battery for identifying potentially successful naval electrical trainees. *J. appl. Psychol.*, 1943, 27, 399–406.

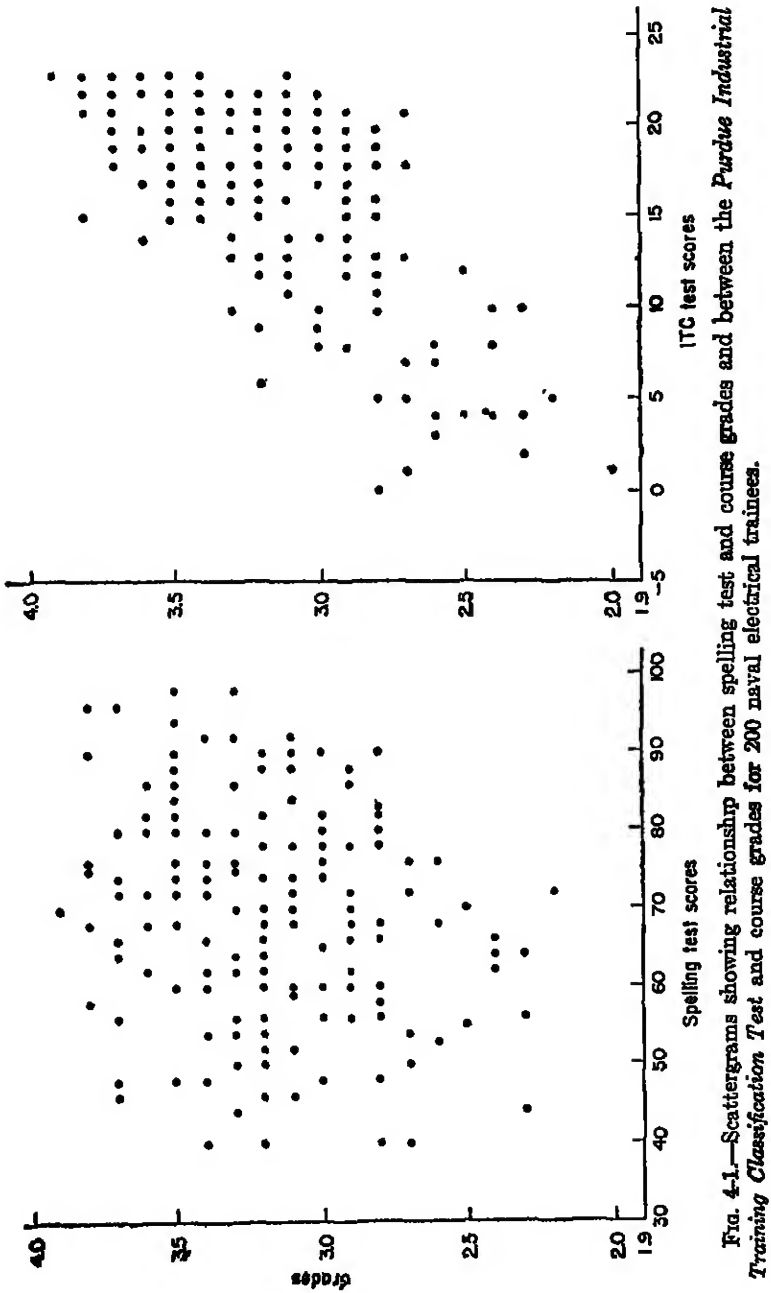


FIG. 4-1.—Scattergrams showing relationship between spelling test and course grades and between the *Purdue Industrial Training Classification Test* and course grades for 200 naval electrical trainees.

describe one variable, in this case scores on a test. The horizontal axis (called the *abscissa*) is used to represent the other variable, in this instance grades in a training program. Each individual is then represented by one dot, the position of the dot being located by that person's test score on the vertical axis and by his grade in the training program on the horizontal axis. When all of the individuals in the study are plotted in this fashion, the resulting scattergram indicates the degree of relationship between the two.

Examples.—The two examples in Fig. 4-1 represent the same group of individuals. They show the relationship of grades in the electrician's training program and each of two tests taken before training began. The scattergram on the left in Fig. 4-1 shows the relationship between spelling-test scores and grades made later in the training program. In contrast, the scattergram at the right shows the relationship between scores on the *Purdue Industrial Training Classification Test* and grades in the training program. It is evident that there is a closer relationship in the case of the latter, since the scatter of dots more nearly approaches a straight line. In the former the pattern more nearly approaches a random scatter or a circle.

Correlation.—These scattergrams are useful when inspected visually. As in the case of the examples in Fig. 4-1, when one test is highly related to the criterion and when the other bears little or no relationship, it is relatively simple to look at the patterns and determine which of the two tests would be most useful for future selection. Sometimes, however, the differences are not so apparent as in the present examples. Statisticians have devised ways of expressing the degree of relationship in terms of a single number called the *coefficient of correlation.*[1] Although the treatment of this concept is beyond the scope of this book, a few words seem in order. The coefficient of correlation describes the degree of relationship between two sets of values. A perfect positive relationship is described by a coefficient of +1.00, and a perfect negative relationship by a coefficient of —1.00. In terms of the examples presented in Fig. 4-1, if a perfect positive correlation

[1] See any elementary textbook in statistics such as LINDQUIST, E. F. *A first course in statistics.* Boston: Houghton Mifflin, 1942. Pp. 53-204.

(+1.00) were to exist between scores on the *Purdue Industrial Training Classification Test* and grades, the dots would all fall into a straight line and it would be possible to predict *exactly* the school grade from the test grade in advance. That is, the trainee having the highest score on the test would also have the highest grade, and the trainee making the lowest score on the test would have the lowest grade, and so on. If, on the other hand, there were a perfect negative relationship (—1.00), it would be possible to predict the grade just as accurately from the test score but the person with the highest test score would have the lowest grade and the person with the lowest test score would have the highest grade and so on. A coefficient of 0.00, indicating no relationship at all, would indicate a random scatter of dots and that an individual having a high test score would be no more likely to have a high grade than would one with a low test score and vice versa. Two variables having a degree of relationship anywhere between these two extremes of no relationship and perfect relationship would yield a coefficient of correlation somewhere between 0.00 and 1.00. In the present examples, the correlation between the spelling test and grades was computed to be 0.26, whereas for the *Purdue Industrial Training Classification Test* it was found to be 0.71. It should also be added that when the coefficient of correlation is known, the value of one variable may be estimated from the other; but as the correlation approaches 1.00, the size of the error will diminish until, if there were a correlation of 1.00, there would be no error. As the correlation approaches 0.00, the error of estimate or prediction increases until, when the correlation is exactly 0.00, the error is as great as it would be if the value of the second variable would always be estimated at the average of the distribution.

Usefulness of Correlation.—The coefficient of correlation is extremely useful in the hands of a skilled technician who is experienced in its use. *Its value is limited largely to the determination of the degree of relationship.* It is extremely difficult to present and interpret to managerial, supervisory, and union personnel not experienced in the use of statistical concepts. Although the scattergram is useful in presenting and interpreting facts about

tests, the person in charge of the program will do well to avoid reference to the coefficient of correlation when making presentations, either verbally or in writing. One or more of the following approaches is ordinarily more useful.

THE METHOD OF AVERAGES

Simple Averages.—As suggested in Chap. II, one of the most understandable and consequently more useful methods for comparing two criterion groups is the method of simple averages. The

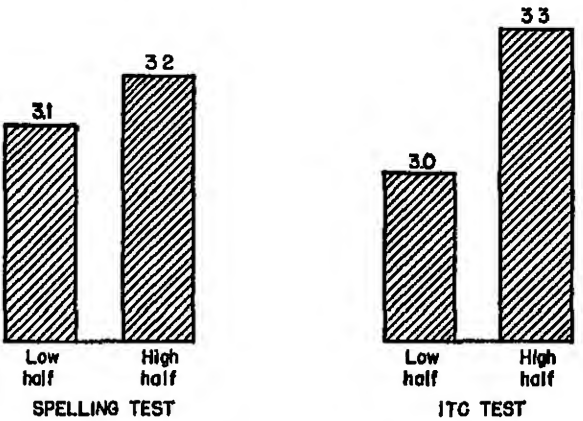


FIG. 4-2.—Average course grades for electrical trainees when divided into high and low halves on two tests.

mean (commonly called the average or arithmetic average) is used. In Fig. 4-2 the same data used above were employed. (1) The 200 men were classified into two groups, the 50 per cent making the highest scores on the spelling test and the 50 per cent making the lowest scores on the same test. (2) The mean (average) grade in the training program for each of the two groups was computed; the pair of bars at the left of Fig. 4-2 show these means of 3.1 and 3.2. The same process was repeated with the *Purdue Industrial Training Classification Test,* and the resulting means of 3.0 and 3.3 are shown at the right of the figure. It is apparent that the obtained difference[1] between the two groups

[1] The question of how great must a difference be before one can be reasonably certain it could not have occurred through chance is discussed in Appendix A.

is greater in the case of the *Purdue Industrial Training Classification Test* than in the case of the spelling test. Actually this is only another way of expressing the same relationship that was illustrated in the scattergrams in Fig. 4-1. Another analysis by means of simple averages could be made by dividing the trainees into two groups on the basis of grades and computing the average score for each test for each group. There are times when the criterion is of such a character that this latter approach is necessary, for example, if employees are rated as either A or B and only

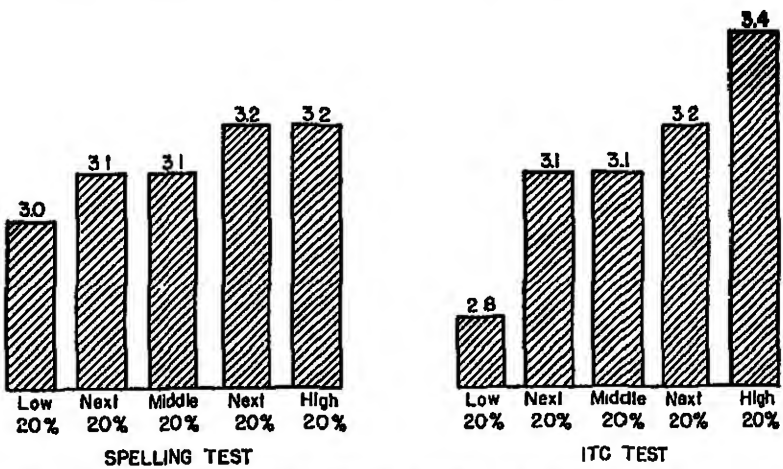


FIG. 4-3.—Successive average course grades of trainees divided into five degrees of excellence on the basis of two different test scores.

two classifications exist or when a comparison is made between the scores of discharged and satisfactory employees.

Successive Averages.—The method of successive averages is identical with the method of simple averages except that the employees are classified into more than two groups for purposes of comparison. In Fig. 4-3, the same facts previously used are employed. However, instead of dividing the trainees into the 50 per cent doing best on the test and the 50 per cent doing poorest on the test, they have been classified into five groups, the best 20 per cent on the test, the poorest 20 per cent on the test, and so on. The average grade of each successive group was computed,

and the two graphs in Fig. 4-3 were prepared. Here again it is evident that the relationship between *Purdue Industrial Training Classification Test* scores and grades is much more marked than is the case with the spelling test. The method is also applicable where employees have been rated into three or more groups. The average test score of each group can be computed.

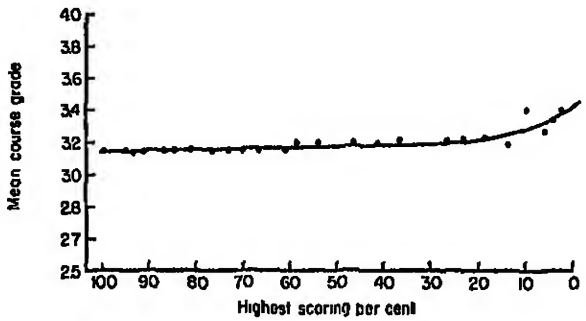


Fig. 4-4.—Curve showing mean course grades of various proportions of the group selected on the basis of highest performance on the spelling test.

Cumulative Averages.—The method of cumulative averages, though occasionally more useful, is sometimes more difficult to explain and interpret to others. In the data that have been used here for illustration, it is known that the whole group had an average grade of slightly more than 3.1. In deciding whether or not to use one of the tests in question for future selection, one might well raise the question, "What would the average have been if only the 50 per cent doing best on the test had been admitted to the program?" Figure 4-4 answers this question with respect to the spelling test. Generally speaking, the smaller the percentage (the best 10 per cent, the best 20 per cent on the test) admitted, the higher the average grade of the group. It will be noted, however, that the percentage admitted would need to be quite small to result in any appreciable improvement in the average of the accepted group. Figure 4-5 presents the same facts for the *Purdue Industrial Training Classification Test.* Here the upward trend of the curve is more marked and more consistent. A comparison of the two figures certainly indicates the superior-

ity of the *Purdue Industrial Training Classification Test* for the selection of these particular trainees. The method is called the method of cumulative averages because in the process of computing the various averages, a cumulative procedure is helpful. The same approach may also be used by allowing the horizontal axis to represent minimum test scores. When this is done, the

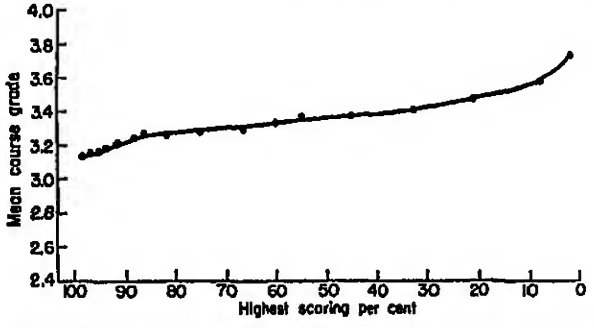


FIG. 4-5.—Curve showing mean course grades of various proportions of the group selected on the basis of highest performance on the *Purdue Industrial Training Classification Test.*

curve answers the question, "What is the average grade of those who made a certain score or higher?"

METHOD OF PERCENTAGES

Simple Percentages.—Frequently data being analyzed are of such a nature that it is either impossible or undesirable to work with averages. In these instances it is possible to work with the percentage who attain a certain standard or meet a certain condition. Using the same facts as in the earlier examples, the question can be asked, "What percentage exceeds the average grade of the total group?" The answers are shown in Fig. 4-6. The trainees were classified into the best half and the poorest half on each test, and the percentage exceeding 3.1 (the over-all average) computed. The same trends that have been noted before still appear, and an evaluation of the two tests can be made in terms of the differences in percentages.

Figure 4-7 expresses these facts in a slightly different fashion. Here the total length of the bar represents 100 per cent of the

group in question. The open section of the bar, like the bars in Fig. 4-6, represents the percentage attaining the standard or meeting the condition; the shaded portion of the bar represents the percentage failing to meet the standard or condition.

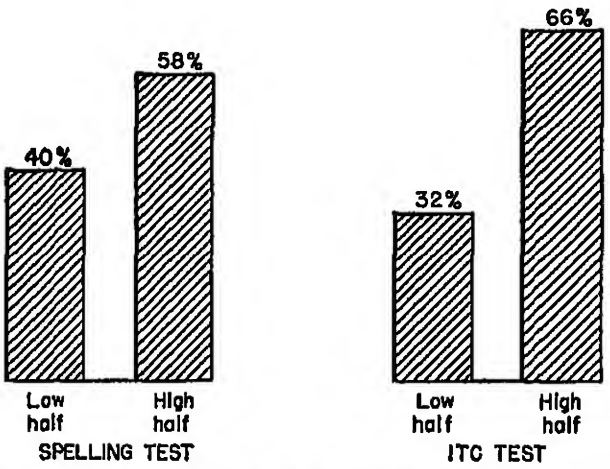


FIG. 4-6.—Percentage of trainees' grades exceeding 3.1 in high and low scoring half of each of two tests.

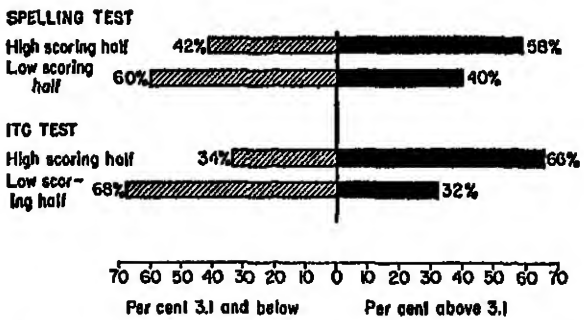


FIG. 4-7.—Percentage of trainees receiving grades above and below critical score when divided into high- and low-scoring half on each of two tests.

Successive Percentages.—The method of successive percentages is actually only a modification of the simple percentage approach. Instead of classifying the individuals into two groups, they are placed in three or more groups and the percentages are computed for each group. Figure 4-8 shows the group classified

into five categories on the basis of spelling-test performance and into five categories on the basis of *Purdue Industrial Training Classification Test* performance. In each instance, the percentage of each subgroup that received a grade of 3.2 or better is shown.

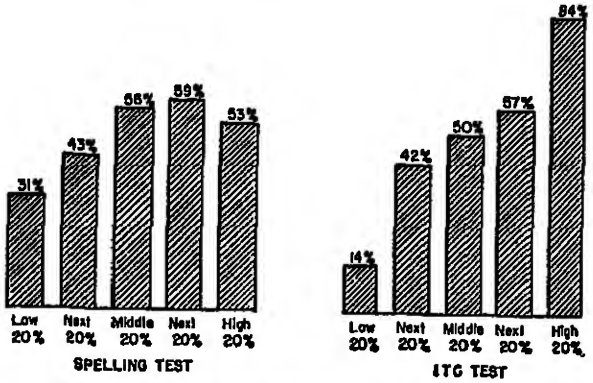


FIG. 4-8.—Percentage exceeding course grade of 3.1 when divided into five degrees of excellence on the basis of each of two different tests.

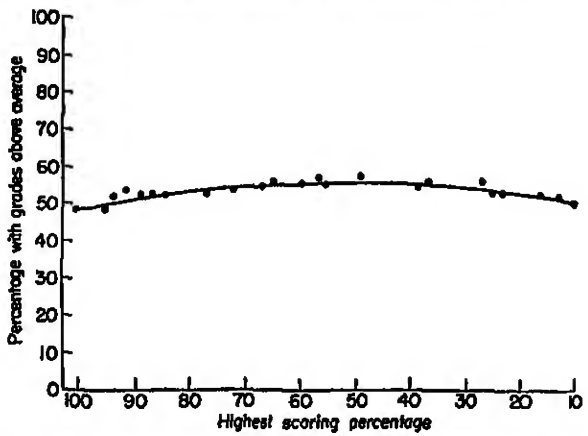


FIG. 4-9.—Percentage exceeding grade of 3.1 in course when varying proportions are selected on the basis of highest performance on spelling test.

The greater relationship in the case of the classification test is again evident.

Cumulative Percentages.—The method of successive percentages indicates the percentage within each test group that attained a given standard. Suppose that the best 30 per cent, the best 50

per cent, or some other proportion had been accepted. What percentage at the various levels would have attained the standard? The method of cumulative percentages answers this question, and Fig. 4-9 shows its application to the spelling test data. Here it will be noted, the proportion receiving grades of 3.2 or better is not materially altered by taking the upper 30 per cent. Figure

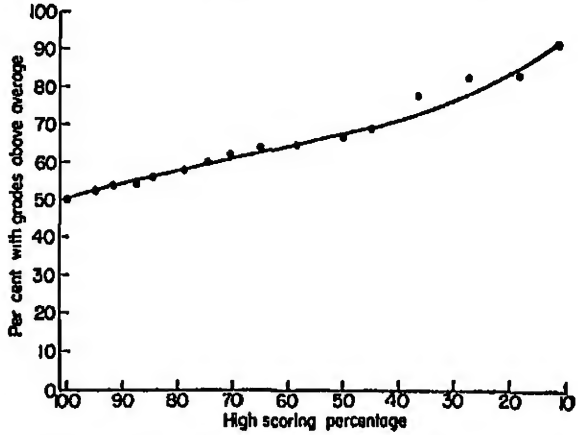


Fig. 4-10.—Percentage exceeding grade of 3.1 in course when varying proportions are selected on the basis of highest performance on the Purdue Industrial Training Classification Test.

4-10, however, treats the *Purdue Industrial Training Classification Test* facts in the same fashion, and it is again clearly apparent that this latter test is superior for this purpose. Generally speaking, the smaller the proportion selected in terms of test superiority, the greater the improvement in the percentage of trainees who receive grades of 3.2 or better.

THE PROFILE METHOD

Meaning of Percentile Scores.—Percentile scores are useful for comparing performance on two or more tests that have different numbers of items and different levels of difficulty. For example, in the case of the two tests used for illustrative purposes in this chapter, one has a maximum score of 23 whereas on the other some trainees score as high as 98. How are comparisons to be made from test to test? What score on the spelling test repre-

sents about the same level of accomplishment as a raw score of 20 on the *Purdue Industrial Training Classification Test?* Percentile scores provide one method for answering these kinds of question. The percentile score equivalent of a raw score value indicates the percentage of individuals in a defined group who scored at that score level or below. For example, if a given raw

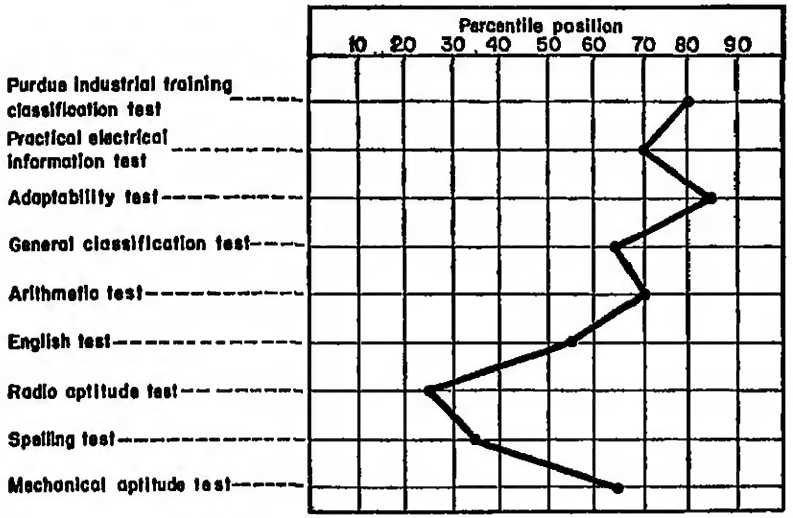


FIG. 4-11.—Percentile graph showing standing of one trainee on each of nine tests.

score falls at the eightieth percentile, it means that 80 per cent of the group made that score or a lower score; the converse is that 20 per cent did better. The fiftieth percentile, then, always defines the mid-point, or the score on the test that was exceeded by half the people. The twenty-fifth percentile and the seventy-fifth percentile indicate the raw-score limits between which the middle 50 per cent of the scores lie.

Individual Profiles.—With this concept, then, it is possible to administer a battery of several tests to a group of employees or applicants and then to convert each individual's raw score on each test to the corresponding percentile value. These percentile scores are then plotted on a chart known as a profile for visual comparison. Figure 4-11 is an example for one electrician trainee

and shows his relative standing on each of the nine tests taken. His standing on both the electrical information test and the arithmetic test is at the seventieth percentile, indicating similar performance on the two tests. The profile indicates that he is above average on all but two of the tests and that he did relatively best on the *Adaptability Test* and poorest on the radio aptitude test.

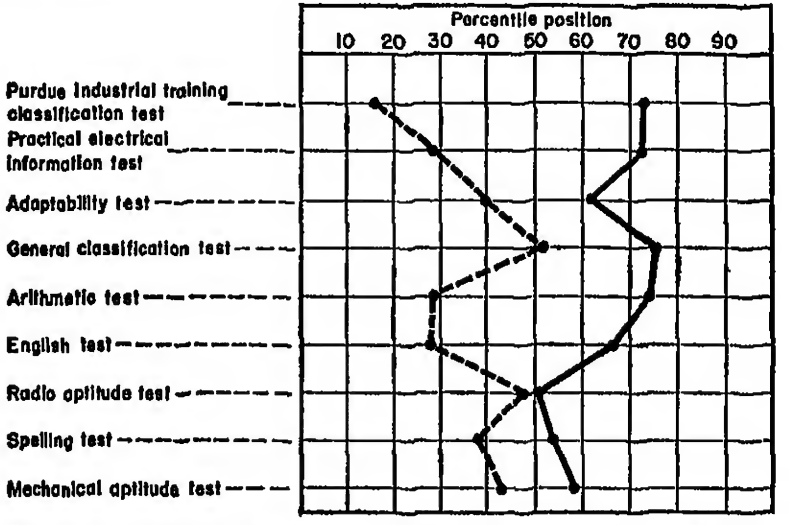


FIG. 4-12.—Percentile graph showing the mean profile of the fourth of the class with highest course grades and the fourth with the lowest course grades.

Group Profiles.—Just as a profile can be prepared for an individual, so can one be prepared to demonstrate the average performance of a subgroup. Frequently it is desirable to plot the profiles of two criterion groups of employees, one that is meeting the standard of job success and one that is not. For example, in the electrician trainee study two criterion groups were identified. The 25 per cent who made the best grades in the school were called "good"; the 25 per cent who made the poorest grades were called "poor." A comparison of the profiles of these groups makes possible the evaluation of the several tests involved. For example, if on a certain test the percentile position of the average score of the good group is no higher than the corresponding percentile

score of the poor group, the test does not discriminate. Figure 4-12 shows the profiles of these two groups. It is apparent that the tests are not equal in the degree to which they discriminate between the criterion groups. The radio aptitude test, for example, shows practically no difference, whereas the *Purdue Industrial Training Classification Test* shows considerable spread. Once such a group profile has been used to indicate what tests in the battery are likely to be useful, it can be used as a standard against which to match individual profiles like the one in Fig. 4-11. It provides a basis upon which to judge the extent to which an individual pattern corresponds to the pattern of the superior group.

SUMMARY

Data from an electricians' training program have been used to illustrate the four basic methods for analyzing and interpreting the relationship between test scores and measures of job success. The methods discussed, the scattergram, the method of averages, the method of percentages, and the profile method, each has minor variations, one of which will satisfy the demands of each specific problem.

CHAPTER V

MENTAL ABILITY TESTS

The Meaning of Mental Ability.—Mental ability or intelligence is more easily measured than it is defined. This may sound strange until one considers the fact that physicists measured electricity long before they were able to understand its nature. Although there is still some disagreement among psychologists regarding the nature of mental ability, there is general agreement that mental ability is reflected in "quickness on the trigger," ability to learn, versatility, and general competence. Some people learn new tasks more quickly than others. Some can figure their way out of problem situations more readily than others. Some are capable only of relatively unchanging, repetitive activity; others are extremely versatile in the sense that they are able to adjust themselves to situations which require the rapid shifting of attention among a variety of different stimuli and the making of certain judgments regarding their relationships. Some people are simply more competent than others in a general way, not alone because they possess specific skills or know certain facts but because they are capable of meeting situations that many times require the manipulation of abstract ideas or concepts. In a sense, mental ability represents a kind of mental "horsepower" rating, and some activities or jobs require more horsepower to get them done than do others.

Primary Mental Abilities.—The research conducted by Thurstone [1] and others indicates that this competency which characterizes the behavior of so-called "more intelligent people" is not a single trait or ability but, more accurately, is a combination of several quite specific abilities. Researchers disagree regarding the extent to which these primary abilities are interrelated; but when

[1] THURSTONE, L. L. *The vectors of mind.* Chicago: University of Chicago Press, 1935.

these traits are taken together without regard to the exent of interrelationship, most experimenters include them in their concept of intelligence.[1] Below is a list of some of the abilities that have been identified by Thurstone and others.

1. *Verbal ability,* reflected in facility with words and language.
2. *Numerical ability,* required in the simple arithmetic operations but not in the more complex reasoning types of situations.
3. *Memory ability,* characterized by the recall of recently learned, rote memory material.
4. *Visualizing ability,* required in the performance of tasks involving space relationships.
5. *Mental fluency,* required in the making of rapid responses or adjustments to abstract tasks.
6. *Perceptual speed,* required in the rapid identification of differences in visual patterns.
7. *Inductive ability,* required in the discovery and application of some rule or principle that is operating in a situation.
8. *Deductive ability,* representing what is most often popularly referred to as reasoning ability.

All present-day tests of mental ability are made up of questions that measure several of these various abilities. The various tests differ primarily in the emphasis that is placed on each and the nature of their organization.

Kinds of Tests.—Some mental ability tests have subparts that are intended to measure some of these specific abilities or combinations of them; others are made up of questions that sample these various abilities in a more or less random fashion. In business and industry the latter has, generally speaking, been most useful, although there are outstanding instances in which the more specific tests have done exceedingly well. One common classification that has been employed by test makers has resulted in the construction of language and nonlanguage tests, the latter

[1] For a comprehensive and critical review of the literature see CATTELL, RAYMOND B. The measurement of adult intelligence. *Psychol. Bull.,* 1943, 40, 153–193.

containing only those questions which are not dependent upon language mastery. Tests may be administered to only one person at a time, or they may be of the group variety with which the number that can be tested at one time is limited only by the testing facilities. The latter is used almost exclusively for personnel placement.

Ability Differences and Their Origin.—It is well known that individuals differ markedly in any single one or combination of the specific abilities listed above. Because the term *intelligence* was adopted before the more recent investigations were made, there has been a widespread belief or, at least, an implication that intelligence or mental ability is innate and that it is relatively independent of the various environmental influences with which an individual comes in contact. In fact, certain psychologists have defined intelligence as just that. It is difficult, however, to consider a concept of this nature which is not measurable, and purely innate abilities do not lend themselves to measurement. An individual may be given verbal or nonverbal tasks to perform, and his ability to perform them can be measured. He has certain amounts of certain abilities *now*. The fact can be verified, but to speculate regarding the relative contribution of heredity and environment to his present status leads to highly theoretical considerations which are extremely difficult to investigate. Although such matters are important from the standpoint of social welfare and educational philosophy, they are quite unimportant in the matter of personnel placement. If a given test, by one or more of the simple techniques presented in Chap. IV, has been shown to be related to some measure of job success, the question of how the particular employee or applicant got that way is unimportant. Many psychologists have adopted the definition that mental ability is that ability or those abilities which mental ability tests measure. This is not a facetious statement and is actually the most meaningful of all definitions. As a matter of fact, the many tests of mental ability yield markedly consistent results. In fact, they are more in agreement than are the people who make them when they start discussing such questions as heredity vs. environment and one vs. many traits. People differ in their abilities to

perform certain tasks in tests and elsewhere; these abilities which they possess *now* are actually the product of environment and heredity. The important fact is that, provided a particular test correlates with success on the job, how an individual got that way is relatively unimportant.

Inequalities in Opportunity.—Some of the more recent researches have indicated that mental ability test scores are markedly affected by the educational and the socioeconomic opportunities that an individual has had. For this reason, the statement is sometimes heard that a given individual should not be penalized in his test score because he has never had the opportunity to develop the skills or abilities tested. Again, it should be considered that if test scores are correlated with job success, how he got that way is unimportant from the standpoint of industrial placement, important as it may be from the social point of view.

The I.Q. and Other Scores.—Everyone has heard of the I.Q. as an index of a person's performance on a mental ability or intelligence test. The I.Q. is the intelligence quotient. It is a quotient representing the ratio of an individual's mental age to his chronological age. It was developed for use with maturing children and actually yields the ratio between a child's mental development and his chronological development. Thus a child of any age, who has matured mentally at the same rate as the average child his age is given a quotient of 100. If his mental development has been more rapid than the average, he has an I.Q. greater than 100; if his mental development has been arrested or has been otherwise slower than the average for his age, he has an I.Q. smaller than 100. More than half of the children of any age group have I.Q.'s between 90 and 110. Since the I.Q. is a ratio, and since the kinds of ability that intelligence tests measure generally cease to improve after late adolescence, the I.Q. is much less meaningful at the adult level than it is during childhood and adolescence. The authors of tests using the I.Q. select some age usually between sixteen and twenty as the maximum chronological age to be employed in computing the ratio, regardless of the individual's true age.

Tests of mental ability that are constructed primarily for adult

use in industry usually employ the raw score, or number of questions answered correctly, in preference to the I.Q. These, in turn, are usually converted to percentile values[1] so that it is possible to say that insofar as the abilities measured by a particular test are concerned, a given individual is in the upper or the lower so many per cent of applicants or present employees.

MENTAL ABILITY AND JOB PLACEMENT

Occupational Groups and Ability Levels.—The mental ability levels of individuals in various occupational groups differ. Harrell and Harrell[2] present data accumulated during the Second World War, and Fig. 5-1 prepared from data in their report lists seventy-five civilian occupations in descending order of average or mean scores on the *Army General Classification Test* of 18,782 soldiers engaged in these occupations prior to service. The shaded bars mark off the test score limits within which approximately two-thirds of the men in each classification scored. Thus, men who were accountants prior to induction averaged 128 points in contrast to men who were teamsters, who averaged 88 points. Two-thirds of the former scored between 116 and 140 whereas two-thirds of the latter scored between 68 and 107. Although certain difficulties are always encountered when job titles are dealt with, nevertheless the figure serves to demonstrate that occupations tend to attract and to hold individuals in a relatively narrow mental ability range. This finding is corroborated in many other studies one of which was conducted by Pond.[3] In an investigation involving 9,025 employees in forty-four different occupations she found that the average or mean mental test score of employees on each occupation correlated 0.77 with estimates of the intelligence required for these occupations. Bills[4] in a study of 780 clerical

[1] See Chap. IV, p. 50.

[2] Harrell, Thomas W., and Harrell, Margaret S. Army general classification test scores for civilian occupations. *J. educ. and Psychol. Meas.*, 1945, 5, 229–239.

[3] Pond, Millicent. Occupations, intelligence, age, and schooling: their relationship and distribution in factory population. *Person. J.*, 1933, 12, 373–382.

[4] Pond, Millicent, and Bills, Marion A. Intelligence and clerical jobs; two studies of relation of test score to job held. *Person. J.*, 1933, 12, 41–50.

workers in an insurance company classified the jobs into eight grade levels ranging from A to H with the A classification representing the lowest level of job. Table V shows the proportion in each mental ability score bracket in the various classification levels. For example, whereas 82 per cent of those scoring zero to 40 on the test were in A or B jobs, only 26 per cent of those scoring 140 or better were on these jobs. Table V shows consistent trends

TABLE V

Test scores	Job-classification levels			
	A and B	C and D	E and F	G and H
0–40	82	9	9	0
41–60	77	16	7	0
61–80	71	24	5	0
81–100	61	27	12	0
101–120	43	34	20	3
121–140	32	32	26	10
141 and above	26	30	35	9

throughout. Bills also conducted a similar study with 123 stenographic employees and obtained essentially identical results.

Tenure and Mental Test Scores.—Studies of this character have been criticized in that they provide no evidence that persons at these respective mental ability levels are necessarily best qualified for the job. Does more or less mental ability than is possessed by the mine-run of those on the job ensure failure in that job? The answer is definitely negative. However, there is evidence that the probability of success is greater if one is within the preferred range. One of the earliest studies along this line was conducted by Pond [1] in which she used the fact of whether or not an individual either terminated or was terminated within a given period of time as the criterion of success on the job. Figure 5-2 shows her findings in four female and five male classifications or job families. With each classification "preferred ranges" of test scores were set up by inspecting the score distributions of those who stayed on the job a stipulated period of time or longer

[1] POND, MILLICENT. Selective placement of mental workers. I. Preliminary studies. *J. Person. Res.*, 1927, 5, 345–368.

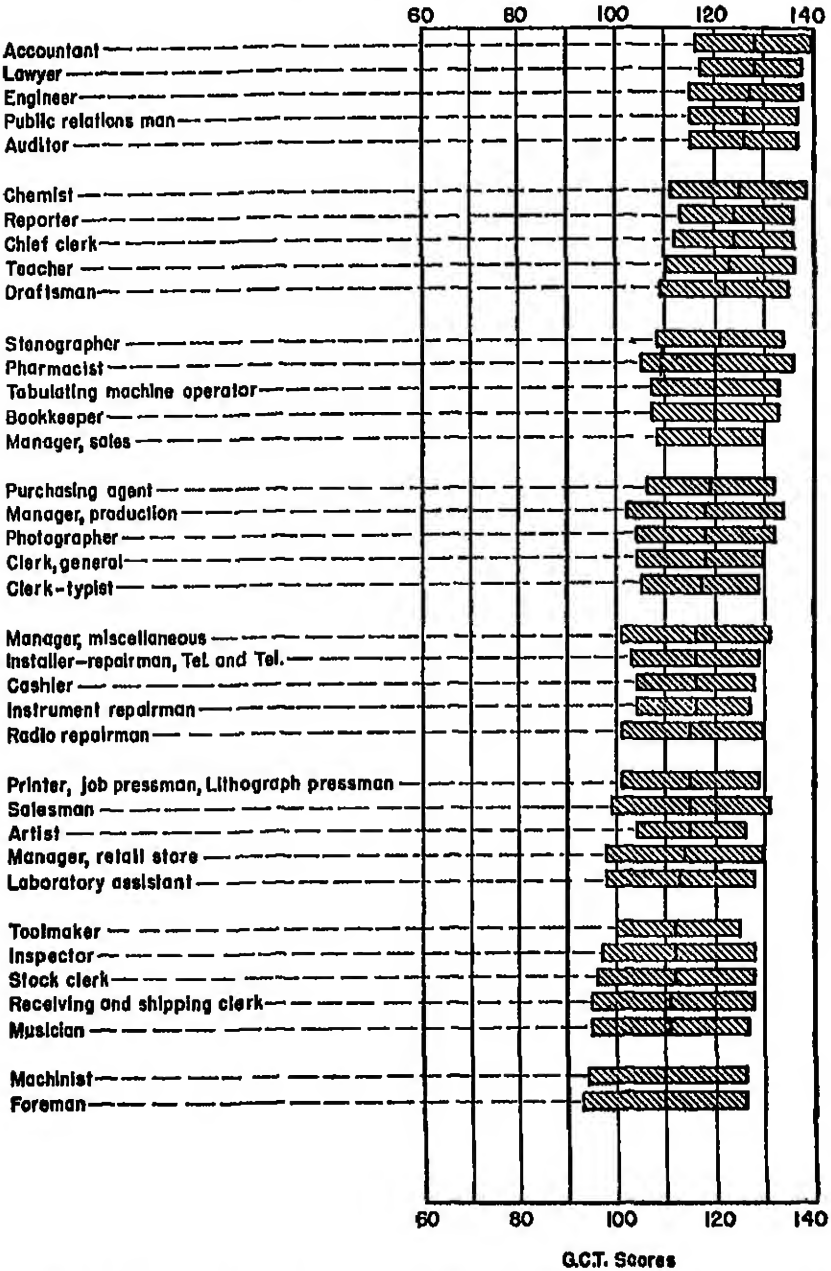
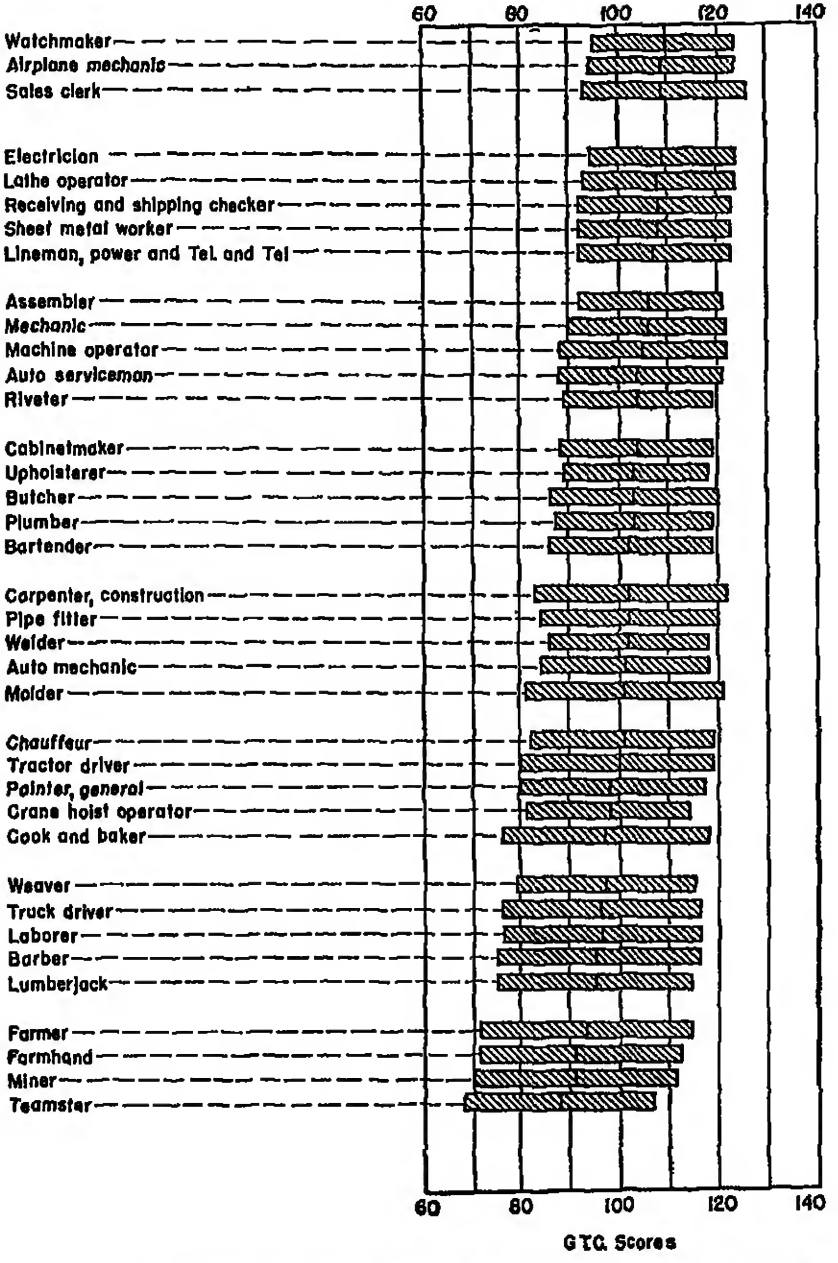


FIG. 5-1.—Average *Army General Classification Test* scores and middle two-civilian occupations. (*From Harrell and Harrell.*)



thirds range of 18,782 Army Air Forces white enlisted men by seventy-five

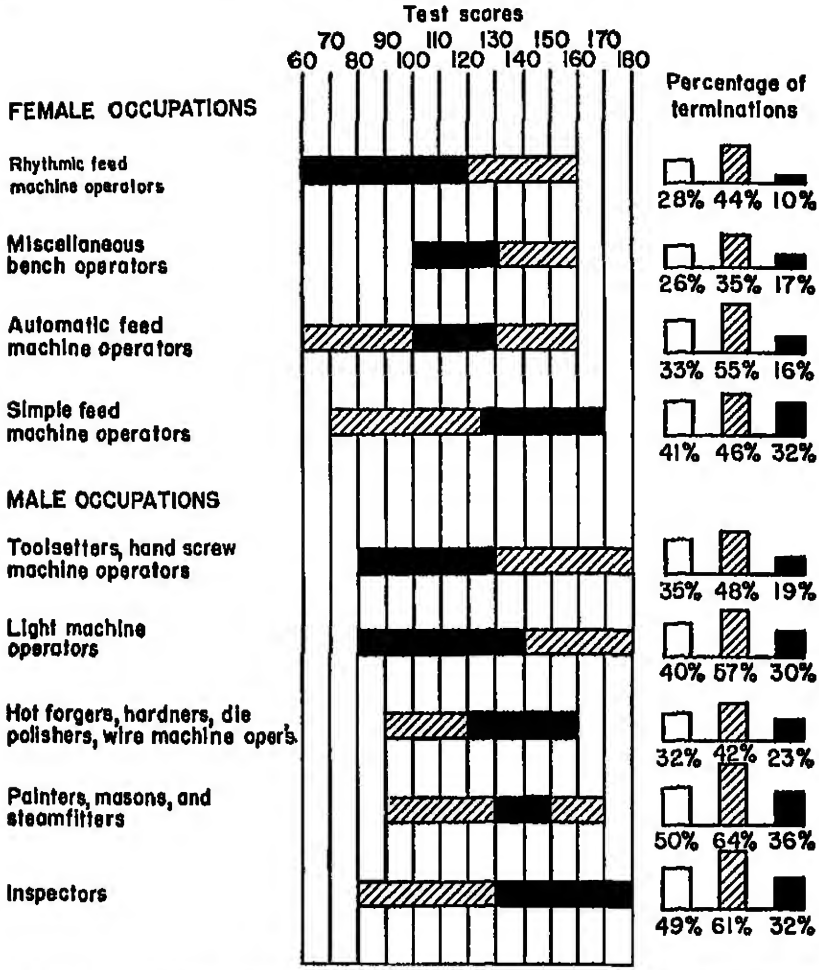


FIG. 5-2.—Preferred score ranges (solid bars) and nonpreferred score ranges (shaded bars) for each of nine job classifications. At the right the vertical bar graphs represent the percentage who terminated in the total group (open bars) among those making nonpreferred scores (shaded bars) and among those who made preferred scores (solid bars). (*From Pond.*)

as compared with that of those who did not remain. The solid bars in the figure indicate the preferred ranges, and the shaded bars the nonpreferred ranges. With both male and female jobs, there is an upper score limit for the lower level jobs, a lower score

limit for the higher level jobs, and both upper and lower score limits for the middle jobs. At the right of the figure, the open bars show the percentage of the total classification group terminated, the shaded bars indicate the percentage of those in the unacceptable range terminated, and the solid bars indicate the per cent in the acceptable score range terminated.

Pond's findings were not limited to these nine classifications by any means. In a study [1] of 3,184 employees in sixty-five job classifications, all treated in this same fashion, she found that whereas 48 per cent of the whole group terminated, only 19 per cent of those in the acceptable ranges terminated. These findings together with similar ones from other studies serve to establish the fact that there is a desirable range of mental ability for each job classification. To place on high-level jobs applicants who have less than the desired amount of mental ability is to place on those jobs persons who cannot deliver what the job demands and so are released, or persons who are made so unhappy in the process of delivering above their normal level that they quit. To place on low-level jobs of a repetitive nature applicants with more mental ability than is required is to place people who are apt to find the job boring, monotonous, or uninteresting. Such people are quite apt to leave for better jobs or, if they do remain, to develop into troublemakers who are eventually terminated.

Personnel departments that have thought of mental ability tests purely in terms of identifying the high scorers have overlooked one of the important values of such tests. A company whose selection program has been based upon such a policy has achieved the most from the tests only if it has a preponderance of high-level jobs. Since job-simplification and production-method trends are continually increasing the proportion of repetitive jobs, the concept of appropriate levels is becoming a more and more useful one.

Measures of Job Satisfaction.—Mental ability is one of the determiners of job satisfaction. The employee who is correctly placed, other things being equal, is more nearly satisfied. One of

[1] POND, MILLICENT. Selective placement of metal workers. II. Development of scales for placement. *J. Person. Res.*, 1927, 5, 405–417.

the very early studies of job satisfaction was carried on by Scott and Hayes[1] before present-day measures of mental ability were as widely used. However, estimates of intelligence were made from educational records on the basis of years retarded or advanced in school. Figure 5-3 shows the percentage of each educational group that was dissatisfied in two different job classifications. In

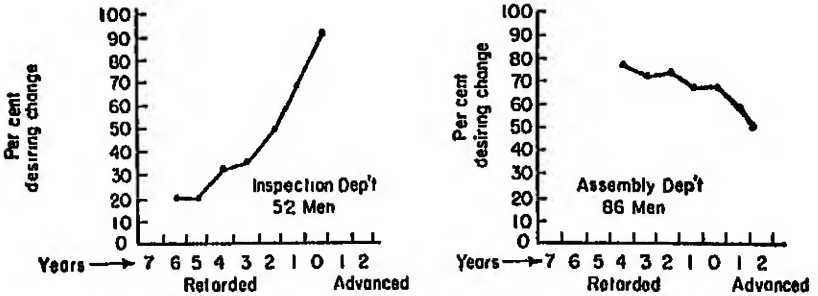


FIG. 5-3.—Percentage of employees on two different jobs expressing desire to change job, classified by years of retardation or acceleration in school. (*From Scott and Hayes.*)

the case of inspection the fewest dissatisfied employees were found among the low-ability (high-retardation) groups, whereas those who were retarded least had a greater tendency to be dissatisfied. This particular inspection job was highly repetitive in nature, requiring a minimum of judgment and related ability. In the case of assemblers, who needed a fair degree of judgment, the picture is quite the reverse. Those who were most frequently dissatisfied were those with the lowest ability, and those who were least frequently dissatisfied were the most able. This is additional evidence that the most able are not necessarily the best on all types of jobs.

Other Criteria of Job Success.—As pointed out in Chap. III, there are many measures of job success. Whether or not an employee is promoted to the next higher job may be one such criterion. In Bills's clerical study[2] previously mentioned, an examination was made of the percentage of employees in various

[1] SCOTT, W. D., AND HAYES, M. H. S. *Science and common sense in working with men.* New York: Ronald, 1921, p. 78.

[2] POND AND BILLS, *op. cit.*

test score brackets who were promoted to higher levels of jobs. Figure 5-4 is based upon her findings. The figure shows that of those scoring 81 or higher on the mental ability test, 75 per cent were promoted to C level jobs or better, whereas of those scoring below that mark, only 36 per cent were promoted. Of those scoring above 100, 56 per cent were promoted to E level jobs or bet-

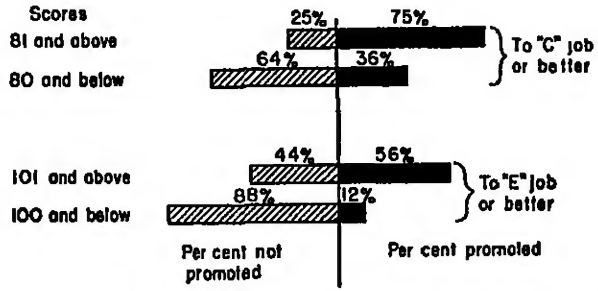


FIG. 5-4.—Percentage of clerical employees scoring above and below two critical scores who were promoted to higher level jobs. (*From Pond and Bills.*)

ter; whereas of those scoring 100 or less, only 12 per cent were promoted.

Wadsworth [1] has also found job differences in intelligence test scores. Like Pond, he established acceptable score ranges, most of which involved minimum critical scores only, because of the character of the jobs in his industry. He made a study of the supervisory ratings of employees who were selected prior to the inauguration of the testing program and compared them with the ratings of employees selected when tests were used to augment the previous selection procedure. His results are presented in Fig. 5-5. Without tests 29 per cent of those hired were rated by supervision as being problem employees, whereas only 5.5 per cent of those selected with the help of tests were so rated. The proportion of employees rated outstanding was at the same time increased from 22 to 33 per cent.

Wadsworth [2] further reports his findings in connection with 793

[1] WADSWORTH, GUY W., JR. Tests prove worth to a utility. *Person. J.*, 1935, 14, 183–187.

[2] WADSWORTH, GUY W., JR. How to pick the men you want. *Person. J.*, 1936, 14, 330–335.

employees in twelve different occupations who were hired with the help of mental ability tests and who were later rated by their supervisors. Figure 5-6 from Wadsworth's data shows the percentage of "outstanding," "satisfactory," and "problem" em-

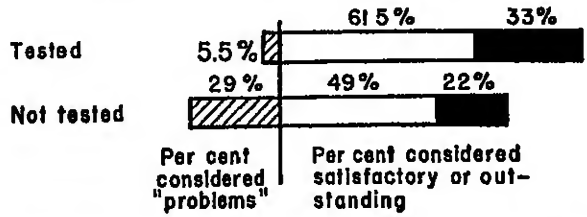


FIG. 5-5.—Percentage considered to be problems (shaded bars), satisfactory (open bars), and outstanding (solid bars), among employees selected by means of tests vs. those selected without tests. (*From Wadsworth*)

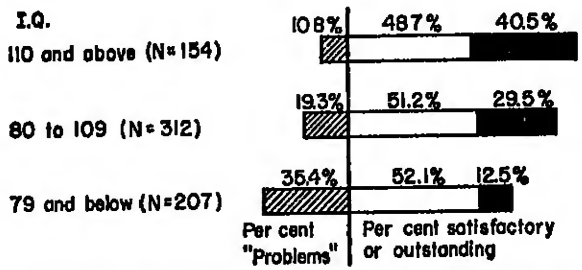


FIG. 5-6.—Percentage considered problems (shaded bars), satisfactory (open bars), and outstanding (solid bars), among employees in three I.Q ranges. (*From Wadsworth.*)

ployees for each of three I.Q. ranges. The proportion of problem employees ranges from 10.8 in the high-intelligence bracket to 35.4 in the low bracket. Whereas the number of employees rated as satisfactory is relatively constant throughout the I.Q. range, the percentage of outstanding employees ranges from 12.5 in the low mental ability group to 40.5 in the high.

SPECIFIC TESTS OF MENTAL ABILITY

There are many tests of mental ability, most of which are good. Generally speaking, these various tests tend to measure the same abilities. Usually the decision of which test to use is made in terms of such factors as (1) the time required for administration,

(2) the orientation provided for the applicant in the directions, and (3) ease of scoring, in addition to the more technical characteristics of the test. It is impossible to mention all available tests[1] here. Those discussed were chosen because of their wide-

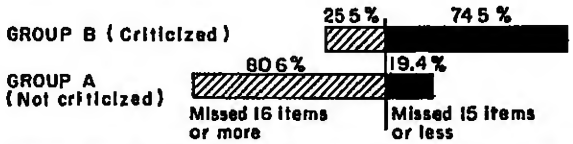


FIG. 5-7.—Percentage of those missing more or less than sixteen items among those managers who were criticized and among those who were not. (*From Stevens and Wonderlic*)

spread use in personnel offices and because validation studies have been published or were otherwise available.

The Otis Self-administering Test of Mental Ability.—The "Otis," as it is called, was used first in schools. Personnel people who needed a test for business and industrial use tended to choose it in preference to other school tests. It has had wide usage, usually with satisfactory results. Stevens and Wonderlic[2] administered it to 160 branch-office managers for a personal-finance company. These men were classified into two groups, group A including those managers who, according to record, had been severely criticized for their methods of systematizing their office procedures, handling details, and generally following work procedures, and group B including those managers who had never been criticized on these bases. Their performance on the Otis was examined in terms of the number of questions missed or omitted. Figure 5-7 shows that of the group with a record of criticism, 80.6 per cent missed sixteen questions or more whereas, of the group not criticized, only 25.5 per cent missed sixteen questions or more.

There have been numerous other validation studies involving the Otis, one of which pertains to supervisors in a textile mill, a report of which appears on page 172 of Chap. XII.

[1] See Appendix C for a comprehensive list of tests and publishers.

[2] STEVENS, SAMUEL N., AND WONDERLIC, ELDON F. The relationship of the number of questions missed on the Otis mental tests and ability to handle office detail. *J. appl. Psychol.*, 1934, 18, 364–368.

The Wonderlic Personnel Test.—The *Personnel Test* is a twelve-minute revision of the Otis test and was designed for employment office use. It has been widely used with excellent results. Typical are those reported by the author [1] himself from a study of the scores made by the outside representatives of a personal-finance company. Figure 5-8 shows these representatives

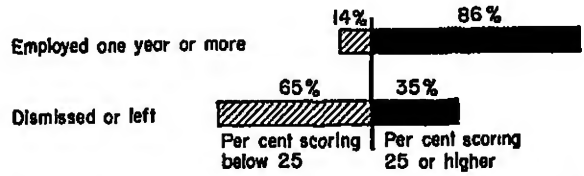


FIG. 5-8.—Percentage of those scoring above or below a critical score on the *Personnel Test* among representatives who stayed on the job a year or more vs. those who did not. (*From Wonderlic.*)

classified as those who stayed with the company a year or more and those who did not. The figure shows the percentage of men in each group who scored 25 and above and the percentage scoring below 25. Whereas only 35 per cent of those who did not stay with the company a full year scored in the higher bracket, 86 per cent of those who did stay with the company a full year or more scored 25 or higher. Wonderlic further reports noticeable differences in the average scores of 100 men selected by supervisors because they had an outstanding record for four years or more and of another group of 100 who had failed to progress. A similar difference was reported between the average scores of a group of outstandingly successful branch managers and another group of managers judged as outstanding failures. The test has been criticized by some because of its short time limit, but Wright and Laing [2] have demonstrated that the differences in results obtained in twelve minutes as compared with twenty-four minutes are negligible.

The Adaptability Test.—The *Adaptability Test* is newer than

[1] WONDERLIC, E. F., AND HOVLAND, CARL I. The personnel test; a restandardized abridgement of the Otis S-A Test for business and industrial use. *J. appl. Psychol.*, 1939, 23, 685-702.

[2] WRIGHT, JAMES H., AND LAING, DONALD M. The time factor in the administration of the Wonderlic personnel test. *J. appl. Psychol.*, 1943, 27, 316-319.

the other two and was designed specifically for personnel placement. It has been widely used since its publication, and much validation information is available. Figure 5-9 shows results obtained with eighty-eight office clerks in a paper mill.[1] These employees were rated by their supervisors as being A (quite satisfactory) or B (not so satisfactory) employees. In the total group

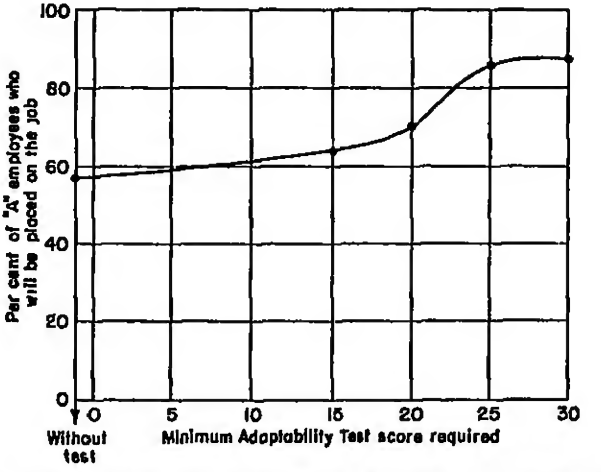


FIG. 5-9.—Graph showing the percentages of clerical employees who were in the A rated (good) group when no test was employed and the percentage who would have been in the A group had successively higher minimum scores on the *Adaptability Test* been used to supplement existing hiring procedures. (*From Tiffin and Lawshe.*)

57 per cent were rated A. However, among those who scored 15 or higher 64 per cent were A's, and of those who scored 30 or higher 88 per cent were A employees.

The test was also administered to a group of electrical trainees[2] prior to training, and the results are presented in Fig. 5-10. Twenty-two per cent of the whole group received a grade of 3.5 or better fifteen weeks later. However, among those scoring 20 or better, 57 per cent made grades in this category. As the figure shows, there is a consistent increase in proportion of

[1] TIFFIN, JOSEPH, AND LAWSHE, C. H., JR. The adaptability test: a fifteen minute mental alertness test for use in personnel allocation. *J. appl. Psychol.*, 1943, 27, 152–163.

[2] *Ibid.*

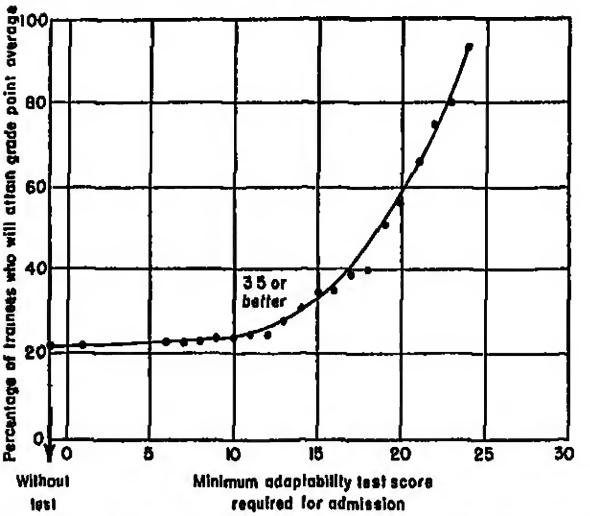


FIG. 5-10.—The percentage of Navy electrical trainees who equaled or exceeded a grade of 35 among who had successively higher minimum *Adaptability Test* scores. (*From Tiffin and Lawshe.*)

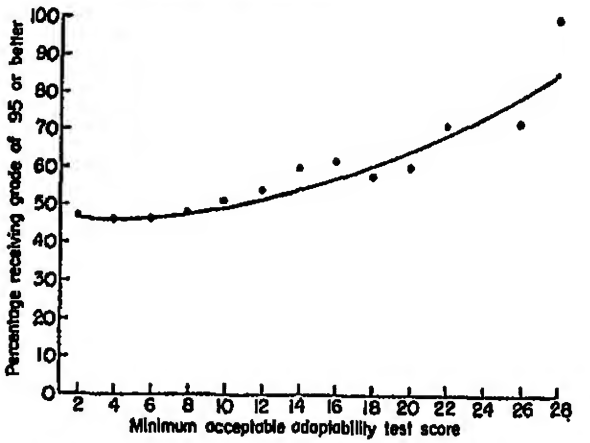


FIG. 5-11.—Percentage receiving a grade of 95 or better in a teletype training program when successively higher minimum *Adaptability Test* scores are applied.

high-graded trainees with corresponding increases in minimum test scores.

In another validation study, 240 employees selected to attend a two-week teletype training program were tested at the time

when training was started. On the basis of objective tests and instructor's judgments they were given percentage grades in the course. Forty-seven per cent of the entire group received grades of 95 per cent or above. Figure 5-11 shows the percentage falling in this grade bracket when successively higher critical scores on the *Adaptability Test* are applied. For example, when only

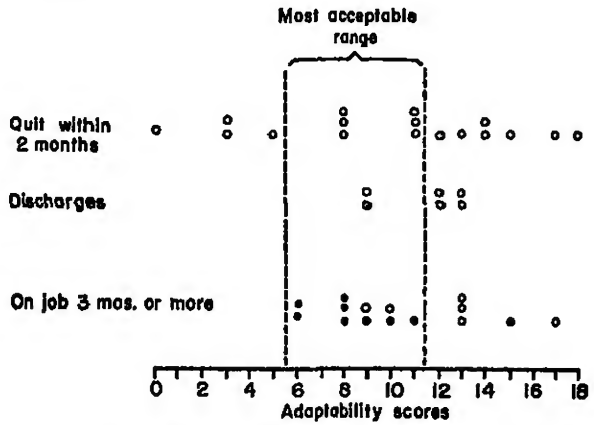


FIG. 5-12.—Spot graph showing acceptable *Adaptability Test* score range for pickler-learners in a steel mill.

those who score 18 or better are considered, approximately 60 per cent received the high grades. The curve shows a consistent increase with successively higher minimum test scores.

Figure 5-12 shows the test scores of thirty-eight applicants who were placed on the job of pickler-learner in a steel mill. Of these men who were tested at the time of hiring, seventeen, or 45 per cent, quit for one reason or another in less than two months, six were discharged for cause, and fifteen were on the job three months or more at the time of the study. The dots in Fig. 5-12 indicate the *Adaptability Test* scores made by the men in each group. As suggested earlier, upper and lower limits were set to establish a score range of 6 to 11. The figure shows that whereas twenty-three, or 60 per cent, of the entire group either quit or were discharged, 75 per cent, or fifteen, of the twenty of those outside the range either were discharged or quit. Those who remained on the job three months or more were rated by

their supervisors, and in Fig. 5-12 those who were rated average or better are represented by solid dots while all others are represented by open dots. Of those within the acceptable range, eight of the ten were rated above average but only one employee outside the range received that rating. When the total picture is considered, it is evident that to get nine employees who would

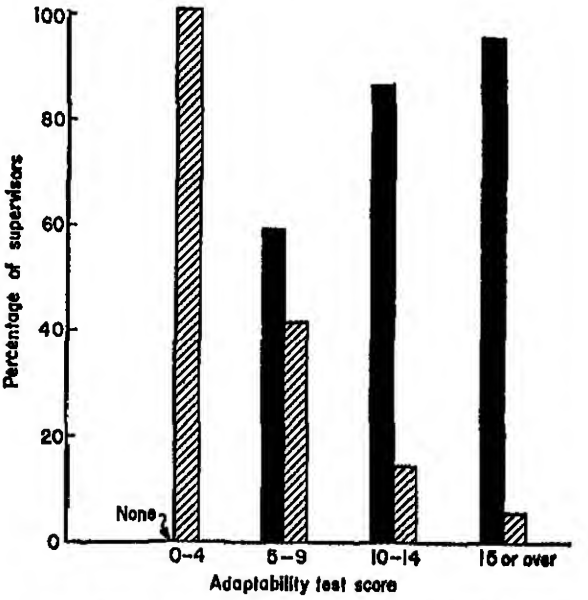


FIG 5-13.—Percentage of supervisors in four *Adaptability Test* score brackets who were on the job six months later (solid bars) and percentage who were not (shaded bars).

be on the job three months later and be considered average or better the company had to hire thirty-eight men representing a ratio of 4.2 to 1. Whereas, had the acceptable range of 6 to 11 been applied eight of the eighteen employed would have fallen in the category representing a ratio of 2.25 to 1.

The importance of mental ability in one supervisory job is demonstrated by Fig. 5-13. Seventy men in a rubber plant were selected by usual means and were upgraded to supervisory positions. After they were selected, each took the *Adaptability Test*, but the test in no way figured in the selection process. Six

TABLE VI.—PLACEMENT RECOMMENDATIONS FOR VARIOUS ADAPTABILITY TEST SCORE BRACKETS

Score on *Adaptability Test* (Form A or B)	Recommendations	Type jobs*
30–35	Should be placed on jobs where there is a real opportunity for the development of originality, independent thinking, and critical judgment. Persons in this bracket are usually very superior on jobs involving verbal or numerical facility. They will learn simple jobs with great rapidity but will soon tire of such jobs and find the mind wandering for new areas to conquer	Potential managers, executives or junior executives, private secretaries
24–29	Should be placed on jobs that require the development of some independent judgment and the opportunity to shift the type of work done from time to time. In general, persons in this bracket learn quickly and easily understand instructions. Some of them, particularly those whose scores are close to 29, have the ability to become as versatile employees as those whose scores are 30 or above	Office clerks, salesmen, apprentices and tradesmen, technicians, supervisors
12–23	Should be placed on jobs that require considerably less independent judgment, adaptability to new situations, and shifting of processes from day to day. This category includes the average or run-of-the-mill persons. They are neither outstandingly high nor outstandingly low in adaptability or mental alertness. They are more adaptable to new situations than persons scoring in the two lower brackets on this test	Setup men, machine specialists, nonautomatic machine operators, store salespeople, storekeepers, standards checkers
6–11	Should be placed on jobs that vary only slightly from time to time and at no time call for exercise of independent judgment. Persons in this bracket are usually below average in ability to understand directions. Job trainers and supervisors will find it necessary to teach the job in a step-by-step manner	Packers, assemblers, inspectors, operators of simple machines which require only routine adjustment
0–5	Should be placed on very simple, repetitive jobs. Persons in this category are at the bottom as far as verbal tasks are concerned They usually have a definite ceiling on their ability to adapt to jobs except those of a simple repetitive nature. While they usually require careful supervision during the learning period, they often become quite expert and self-sufficient on simple jobs where the job does not change	Simple packers, simple assemblers, fixed gauge inspectors, operators of machines that require practically no adjustment or specific care by the operator

* These jobs are presented only as a rough indication of possible employee allocation. Job descriptions and job demands vary so greatly among companies that it is impossible to give a universally applicable list of typical jobs for the different *Adaptability Test* score brackets. (*From Tiffin and Lawshe.*)

months later a study of the group was made, and it was discovered that many were no longer in the supervisory jobs. Some had quit or had been discharged, and some had been demoted to their previous jobs. All of those who were originally tested were divided into four groups according to test score, 0 to 4, 5 to 9, 10 to 14, and 15 or over. As the figure shows, 100 per cent of those scoring in the 0 to 4 bracket were no longer on the job whereas only 5 per cent of those scoring 15 or more were no longer on the job. A systematic trend operates throughout the four score groups. It is certainly evident that the higher men score, the more likely they are to be supervisors in this plant six months later.

The authors[1] of the test have related broad job level descriptions to test score ranges as a guide to the test user. These are reproduced in Table VI but are not intended to be a substitute for validation studies which should be carried on in each plant or business. Job titles are frequently misleading; and the differences in job requirements are so great that even good job descriptions do not always reveal them.

SUMMARY

Mental ability, which is most accurately defined as that ability which the tests measure, is a combination of numerous more specific abilities. Employees who possess these abilities to a greater extent than others tend to be "quick on the trigger," to learn more easily, to be capable of versatility, and to be generally more competent. While performance on these tests reflects the interaction of heredity and environment, the proportional contribution of each is relatively unimportant in personnel placement.

There is a desirable mental ability range for each job classification and particularly for the lower level jobs, upper limit critical scores are just as important as lower limit scores. Validation facts for three of the leading industrial tests are presented.

[1] *Ibid.*

CHAPTER VI

TEMPERAMENT AND PERSONALITY TESTS

The previous chapter has dealt with mental ability and its relationship to job success. However, there has been no intent to imply that the simple possession of the amount or degree of mental ability which a given job requires ensures success on that job. There are other aspects of personality that help to determine the degree of success, not the least important of which is good mental health plus the kind of temperamental pattern that is most easily adapted to the specific job in question. The present chapter deals with this particular phase of personality and its measurement.

THE NATURE OF TEMPERAMENT

Terminology.—Inadequate definition of terms has helped to confuse those who are interested in testing. For example, a common and generally accepted definition of personality is "the sum total of an individual's mental, emotional, or temperamental make-up." In other words, an individual's personality is his total being. This seems simple enough until one encounters a test called a "personality" test. In reality, any test measures some phase of personality, and the term *temperament test* is more meaningful, since such instruments do not measure one's total personality but actually measure one aspect of it.

What Temperament Is.—Psychologists have defined temperament in different ways. The most useful definition for the personnel or industrial relations man is the one that considers temperament as one's behavior tendency. Some people have a tendency to behave in a domineering fashion; some have argumentative tendencies; some are retiring; and so on. These

tendencies can be thought of as descriptive of their temperaments. One person is described as being hot-headed; one as sensitive; and still another as being quite excitable. Theories have been advanced regarding body chemistry as a determinant of temperament, and studies have indicated certain relationships between glandular function and temperament. However, the personnel man can best take the same position regarding temperament as it has been suggested that he take relative to mental ability, namely, that how an individual got that way is a secondary matter. From the point of view of personnel placement it is unimportant, important as it may be from the point of view of social and educational philosophy and clinical psychology. One's temperament is his behavior tendency, his tendency to act in certain characteristic ways, not at any given instant but over extended periods of time.

How Habits and Traits Are Grouped.—Each person in the process of growing, developing, or getting older acquires specific habits of action. Specific as these habits are when considered individually, they tend to group themselves together in clusters. So the individual who is characterized as being shy and who frequently crosses the street to avoid meeting another is most probably apt to retire to an inconspicuous corner at a party or, indeed, not to attend the party at all. Habits or traits of this kind tend to cluster together. Note that the word tend is used; these traits are not always associated with each other. But habits and behavior traits are associated frequently enough so that psychiatrists and others speak of components of temperament. Components of temperament are the larger, more general clusters or combinations of specific habits and behavior traits. For example, selfishness is a component of temperament. Actually, the word *selfishness* merely describes dozens or hundreds of specific habits or traits that an individual might possess. The more such habits and traits he has, the more selfish he is considered to be. Yet selfishness must be thought of only as a tendency that, although present in everyone, is more prevalent in some people than in others.

As a general rule, an individual whose temperamental tendencies are extreme or whose tendencies are in conflict to the

extent that tension, strain, or anxiety is characteristic is said to be maladjusted. His personality is not well integrated; and other things being equal, he will be less satisfactory as an employee.

The Components of Temperament.—There are various theories of temperament, and each one recognizes or emphasizes various combinations of components. Humm's adaptation [1] of Rosanoff's theory [2] recognizes seven patterns of complexes or tendencies. Although these tendencies are present in all personalities, it is the difference in degree or amount that is associated with differences in behavior. These seven components [3] together with their prime tendencies and associated traits follow:

Component	*Prime Tendency*	*Some Associated Traits*
Normal	Self-control	Nervous stability, self-improvement, conservatism, self-direction
Hysteroid	Selfishness	Drive toward advantage and profit, self-preservation, egocentricity
Manic	Excitability	Drive, alertness, cheerfulness, sociability, wide interests
Depressive	Depression	Sadness, worry, caution, retardation
Autistic	Daydreams	Visual imagery, shyness, seclusiveness, sensitiveness
Paranoid	Fixed-ideas	Egotism, durability of opinion, argumentativeness, rationalization
Epileptoid	Project-making	Planfulness, meticulousness, inspiration to achievement

Guilford,[4] while recognizing most of the Humm-Wadsworth components, fractionates some of them and adds others. His list follows:

[1] HUMM, DONCASTER G. Personality and adjustment. *J. Psychol.*, 1942, 13, 109–134.

[2] ROSANOFF, A. S. *Manual of psychiatry.* (6th Ed.) New York: Wiley, 1927.

[3] HUMM, DONCASTER G. *Personnel evaluation method.* Los Angeles: Doncaster G. Humm Personnel Service, 1945.

[4] The component descriptions presented here have been adapted from material on the *Guilford-Martin temperament profile chart.* Beverly Hills, Calif.: The Sheridan Supply Co.

Component	*High Degree*	*Low Degree*
Social introversion-extroversion	Sociability, tendency to seek social contacts and to enjoy the company of others	Shyness, tendency to withdraw from social situations and to be seclusive
Thinking introversion-extroversion	Lack of introspection and an extrovertive orientation to the thinking process	Inclination to meditative thinking, philosophizing, analyzing one's self and others
Rhathymia	Tendency to be happy-go-lucky or carefree, lively and impulsive	Inhibited disposition and overcontrol of impulses
Depression	Freedom from depression, cheerful, optimistic	Chronically depressed mood, including feelings of unworthiness and guilt
Cycloid disposition	Stable, emotional reactions, freedom from cycloid tendencies	Cycloid tendencies as shown in strong emotional reactions, fluctuation in mood, tendency toward flightiness and instability
General activity	Tendency to engage in vigorous overt action	Tendency to inertness and a disinclination for motor activity
Ascendance-submission	Social leadership	Social possessiveness
Masculinity-femininity	Masculinity of emotional and temperamental make-up	Femininity of emotional and temperamental make-up
Inferiority feelings	Self-confidence	Lack of confidence, undervaluation of one's self, and feelings of inadequacy and inferiority
Nervousness	Tendency to be calm, unruffled, and relaxed	Jumpiness, jitteriness, and tendency to be easily distracted, irritated, and annoyed

Component	*High Degree*	*Low Degree*
Objectivity	Tendency to view one's self and surroundings dispassionately	Tendency to be hypersensitive and to take everything personally
Cooperativeness	Willingness to accept conditions with tolerant attitude	Overcriticalness of people and things
Agreeableness	Lack of quarrelsomeness and domineering qualities	Belligerent and domineering attitude and overreadiness to fight over trifles

While many temperament scales designed to measure one or more components have been developed, generally speaking the components measured appear in one or both of the above lists.

THEORY OF TEMPERAMENT MEASUREMENT

The Basis of Measurement.—Although test makers and theorists often disagree in their concepts of temperament and in whether or not they accept or reject certain components, they do agree in one basic respect; that is, an individual's present and previous behavior is the best key to his future behavior. Consequently the questionnaire approach is utilized. Questions are employed that are designed to determine how a person says he has behaved in past situations or how he feels or thinks about specific situations now. Hence, to the extent that the questions selected tend to sample traits or habits identified with the component in question, a quantitative statement of that tendency is obtained.

How Items Are Selected.—In the preparation of temperament scales, authors first prepare a comprehensive list of questions intended to measure the component in question. The questions included are selected in terms of clinical experience or of a logical analysis of the component. These questions then receive a tryout with criterion groups. The groups may be composed of clinically chosen individuals; one of them may be composed of institutionalized cases; or they may be made up of persons who

have been judged by supervisors or others to be extreme insofar as the tendency is concerned. An item analysis is then made to determine what proportion of the high-criterion group and what proportion of the low-criterion group gave a particular response to the question. To the extent that the question has been answered differently by the two groups, the item is said to discriminate and is retained as a good item; to the extent that both groups tend to give the same answer, the item is rejected. As a result of this procedure,[1] the best items are retained for inclusion in the final form of the scale. Authors of these scales have pointed out that rarely can one predict with absolute certainty how a given item will be responded to or how well it will discriminate. In an effort to mask or withhold their true attitudes or feelings, individuals taking the tests sometimes give answers opposite from those expected, which answers nevertheless may be equally indicative in terms of the statistics.

Kinds of Tests.—Temperament tests are occasionally designed to measure one component. Allport's *Ascendance-submission Test* is an example. More frequently, however, the scales are multiphasic in the sense that they are designed to measure several components of temperament at once. In the latter case, a profile [2] is generally used to interpret the test for a single individual. Frequently, one's score on a given component is not so important as is that score in relation to his score on one or more other components.

Accuracy of Report.—Questionnaires of the "yes or no" variety have been criticized because the person taking the test is at liberty to answer either way in spite of how he really feels at the time or how he acts under certain circumstances. It is further alleged that not all individuals are capable of reporting accurately, even though their intent is good. Humm has attacked these criticisms in two ways, first by providing an industrial standardization of his test; in other words, his norms have been established on persons who were applying for a job and all of whom found it to their advantage to present what seemed to

[1] See Chap. XIII, pp. 183–188.
[2] See Chap. IV, p. 51.

be the best picture of themselves. His second approach is through an evaluation of the number of "no" answers given and an analysis of the balance among the various components. Persons who present inaccurate pictures of themselves tend to "underreport" or "overreport." An extremely negativistic person accumulates more than the expected number of no's, while the extremely suggestible person tends to accumulate fewer than the expected number of no's. On the basis of his research he first established an acceptable range for the "no" count, and any record was considered acceptable or nonacceptable in terms of whether or not it met this criterion. He has since developed techniques for correcting scores for records having unacceptable "no" counts so that practically no records are now eliminated for this reason.

In terms of practical application, however, these are theoretical questions to be left to the test maker and the research man. If a test discriminates between successful or unsuccessful employees by some techniques such as those outlined in Chap. II, then whether applicants reported accurately or not becomes a matter of secondary importance.

SPECIFIC EXPERIENCES WITH TEMPERAMENT TESTS

The Humm-Wadsworth Temperament Scale.—The Humm-Wadsworth [1] is one of the more widely used in industry of all temperament tests. It consists of over three hundred questions of the "yes or no" variety and is scored in terms of the seven components listed earlier. The scale was originally validated by administering the initial list of questions to seven pairs of criterion groups representing the two extremes of each of the components. As a result, the discriminating questions were identified; and although a testee still answers all the questions, only about half of them are scored. For each person taking the test a profile is provided, and his performance is interpreted in terms of his total picture on the components listed on page 77.

An example of the applicability of the scale has been provided

[1] HUMM, DONCASTER G., AND WADSWORTH, GUY W., JR. The Humm-Wadsworth temperament scale. *Amer. J. Psychiat.*, 1935, 92, 163–200.

by Dorcus[1] who administered the scale to fifty industrial employees, divided into two criterion groups, not on the basis of production but on the "basis of the fact that they showed signs of maladjustment and discontent or were problem employees." Without knowledge of how the fifty employees were classified, he designated each individual as satisfactory or unsatisfactory on the basis of the test profile. Figure 6-1 shows the results obtained. Of those classified as satisfactory on the basis of the test, 75 per cent had been previously rated as satisfactory em-

FIG. 6-1.—Proportions of those considered satisfactory and unsatisfactory on the Humm test who were rated satisfactory and unsatisfactory by supervisors. (From *Dorcus.*)

ployees whereas only 25 per cent had been considered as unsatisfactory. Of those classified as unsatisfactory on the basis of the test, 67 per cent had previously been considered as unsatisfactory whereas only 33 per cent had been rated satisfactory. In other words, Dorcus, using the test, properly identified 75 per cent of the satisfactory employees and 67 per cent of the unsatisfactory ones. Humm[2] has criticized this study on two counts. First he says that standardized test conditions were not observed and that the employees in this instance were reassured too much that "the test in no way will affect your status with the company." This, he says, has resulted in relieving the testees of certain tensions and fears, which caused them to respond differently from the case with the standardizing groups. He maintains that special training is necessary and that had the proper procedures and interpretations been utilized the results would have been

[1] DORCUS, ROY M. A brief study of the Humm-Wadsworth temperament scale and the Guilford-Martin personnel inventory in an industrial situation. *J. appl. Psychol.*, 1944, 28, 302–307.

[2] HUMM, DONCASTER G. Discussion of Dorcus' study of the Humm-Wadsworth temperament scale. *J. appl. Psychol.*, 1944, 28, 527–529.

in greater agreement. Humm also expresses his opinion that this is not a validity study. He says that it is a study of the applicability of the scale to an industrial situation. Validity, he implies, is determined by how well the scale identifies those who possess strong or weak tendencies insofar as the various components are concerned. Humm [1] has conducted such studies and has reported coefficients of correlation in the high 90's between profile interpretations and clinical and case history studies.

Guilford's Inventories.—Three different scales are used to measure Guilford's thirteen components listed on pages 78 and 79. These are his *Inventory of Factors STDCR,* intended to measure social introversion-extroversion, thinking introversion-extroversion, rhathymia, depression, and cycloid depression; the *Inventory of Factors GAMIN,* designed to measure general activity, ascendance-submission, masculinity-femininity, inferiority of feelings, and nervousness; and the *Personnel Inventory* I, intended to measure objectivity, cooperativeness, and agreeableness. Of the three, the last is receiving greatest acceptance in industry. According to Martin [2] it measures three subfactors of the paranoid component and is most useful in identifying the potential troublemaker. He presents evidence to support his statement from two industrial situations.

Fifty-one employees in an aircraft-parts manufacturing plant who were rated by supervision as satisfactory or unsatisfactory in terms of adjustment were given the scale. In another study forty-three textile mill employees were rated and tested in the same fashion. In each instance the test properly placed between 70 and 75 per cent of the employees.

Dorcus [3] used the *Personnel Inventory* I at the same time he used the *Humm-Wadsworth Temperament Scale,* and his findings are presented in Fig. 6-2. Of those classified as satisfactory by the test, 73 per cent were rated satisfactory; and of those

[1] HUMM, DONCASTER G., AND HUMM, K A Validity of the Humm-Wadsworth Temperament Scale. *J. Psychol.,* 1944, 18, 55–64.

[2] MARTIN, HOWARD G. Locating the troublemaker with the Guilford-Martin Personnel Inventory. *J. appl. Psychol.,* 1944, 28, 461–467.

[3] DORCUS, *op cit.*

classified as unsatisfactory by the test, 68 per cent were rated unsatisfactory.

While the authors of the scales believe that the cooperativeness component is the most important of the three, they believe that all three are important. Guilford [1] cites one study in which a group of employees were classified as being above average on two or three of the components or below average on two or three of the components. His results are presented in

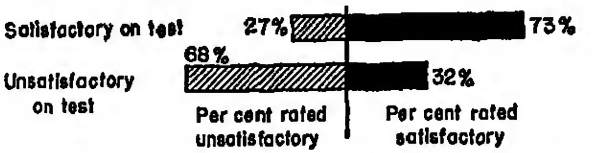


FIG. 6-2.—Proportions of those considered satisfactory and unsatisfactory on the *Personnel Inventory* I who were rated satisfactory and unsatisfactory by supervisors. (*From Dorcus.*)

Fig. 6-3. Of those rated as being satisfactorily adjusted on the job, 66 per cent were above average on two or three of the components; of those rated as being adjusted in an unsatisfactory manner, 73 per cent were below average on two or three of the components.

Although the inventory is relatively new, the authors have collected enough tryout data [2] to make the following claims for it:

1. Approximately 85 per cent of the employees rated as malcontents by management are found to have received a raw score below 60 on the cooperativeness trait.
2. A number of executives who have taken the test have been found to be well above the average on both objectivity and cooperativeness but somewhat below average on agreeableness. In other words, the lack of dominating qualities in an employee is an important factor in making him a contented, peaceable worker; but in the case of an executive,

[1] *Manual of directions and norms.* Beverly Hills, Calif.: Sheridan Supply Co.

[2] *Some information about the Guilford-Martin Personnel Inventory* I. Beverly Hills, Calif.: Sheridan Supply Co. (Mimeo.)

dominating qualities may be an important contributing factor to his supervisory success.

3. We have found some few cases where the testee fails to answer questions truthfully. This seems to be the exception rather than the rule and to occur most often if they do not answer the questions as they believe management wants them to answer. The more intelligent the person is, the better can he falsify consistently if he is so inclined.

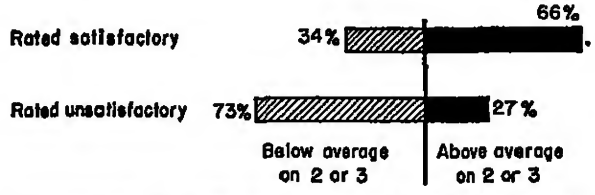


FIG. 6-3.—Proportions of those employees rated satisfactory and unsatisfactory who were below average or above average on two or three of the components in the *Personnel Inventory* I.

Experience with Other Scales.—As stated earlier, there are many scales and inventories designed to measure one or more temperament components. The use of many of these has been confined to the psychological clinic and the laboratory, and reports of their use in business and industry have been scanty. Bernreuter's *Personality Inventory* has been widely used in the selection of salesmen. Much of this work has been done by consulting companies and reports have not been available in published form. Rosenstein [1] reports that "four factors described and considered by this questionnaire have been found of value in the selection of salesmen, extroversion, dominance, self-confidence, and social independence or self-sufficiency." Chapter XI presents some facts relative to the value of the Bernreuter in combination with other measuring devices.

Schultz [2] administered the Beckman *Revision of the A-S Test* and the Root *Introversion-extroversion Test* to a group of 259

[1] ROSENSTEIN, J. L. *The scientific selection of salesmen.* New York: McGraw-Hill, 1944. 162 pp.

[2] SCHULTZ, RICHARD S. Standardized tests and statistical procedures in selection of life insurance sales personnel. *J. appl. Psychol.*, 1936, 20, 553–566.

insurance salesmen. His results presented in Fig. 11-1 and discussed in Chap. XI are pertinent here. Among the poorest producers, 53 per cent were considered unacceptable on the test; among the best producers, only 32 per cent were considered unacceptable on the test.

One of the newer scales and one that is based upon a slightly different approach is Jurgensen's *Classification Inventory*. The scale differs in two important respects. (1) Using a new type of item, the author has endeavored to include only those which research has indicated are at least subject to falsification by the testee, and (2) he recommends the scoring of the scale by occupations rather than by components. This approach is somewhat revolutionary, and little validity information is at present available. However, he does report a coefficient of correlation of 0.80 between test scores and criterion measures in the case of forty salesmen. Further research will be needed to establish the validity of the scale.

SUMMARY

While the term temperament is not universally defined, it is used here to denote behavior tendency, that is, the tendency of an individual to behave in characteristic ways. Behavior acts tend to cluster into patterns usually called *components*. Some temperament scales are designed to measure a single component, but most scales are intended to measure several. The problem of intentional and unintentional falsification on inventory types of scales is a real one but is relatively unimportant when the same validation procedures discussed earlier are employed.

CHAPTER VII

INTEREST AND PREFERENCE TESTS

Some people enjoy human contacts; some enjoy manipulative activities; and still others prefer to engage in verbal activities. What an individual likes, in addition to his mental ability and his temperament pattern, is one factor associated with occupational adjustment and success. This chapter deals with the measurement of interests and preferences.

THE MEASUREMENT OF INTERESTS

Types of Items.—Interest and preference tests utilize a number of different types of items or questions, but a discussion of two major ones will suffice as illustrations. The preference item is one containing two or more responses. Each response represents an activity or an object, and the individual is asked to indicate the one he likes most, sometimes the one he likes least, sometimes the one he likes second best, and so on. The respondent must make a choice between activities or objects. In the other type of item, he is given the name of an activity or an object and is asked to state whether he likes it, dislikes it, or is indifferent to it. Both types of items have produced excellent results, and no effort is made here to demonstrate the superiority of one over the other.

Basic Scoring System.—There are two basic scoring systems presently in use, each of which has a counterpart in temperament testing. The first is the scoring of the test by specific occupations, in which a different key is prepared for each occupation. Thus, answering "dislike" to a particular item might contribute to a high score on the lawyer key and to a low score on the life-insurance-salesman key. Strong has standardized his *Vocational Interest Blank for Men* by this method, and at present there are available twenty-seven different occupational scoring keys.

The other approach utilizes the component approach. Thus, the items indicating interest in mechanical activities are grouped together, those indicative of scientific interest are grouped together, and so on. The applicant receives several different scores on the same test, one for each of the components measured by it. For each person, a profile is prepared and an evaluation of his relative interests in different fields can be determined. For example, he may be in the upper 10 per cent of the general population in the strength of his mechanical interest, whereas he may be in the lower fourth of the same population insofar as interest in literary activities is concerned. This approach is used in the *Kuder Preference Record.* Both approaches have their advantages and their disadvantages in terms of the specificity of their results.

How Items Are Validated.—Items in tests or scales of this character are validated in a fashion similar to the procedure discussed earlier in conjunction with temperament testing. This is particularly true in the preparation of occupational keys. In the process of preparing a key for the job of accountant, for example, the scale is administered to a large group of accountants who have been engaged in that occupation for a minimum stipulated period of time. Item counts are made, and the proportion of accountants responding in a given way to a specific question is compared with the proportion of nonaccountants or the proportion of a general population responding in the same way. To the extent that accountants tend to answer the item in the same fashion as the other group, the item is given a weight of zero for scoring interest in that *particular occupation;* it may be quite important in another occupation. Likewise, to the extent that accountants appear unique in that they tend to answer differently than the mine-run of people, the item is given a weight or value. The range of these values varies with the several tests, but the greater the difference, that is, the more unique accountants are, the greater the weight. Once keys are prepared, one simply adds the various weights assigned to the responses given by a single individual to obtain a total score.

The validation of items for inclusion in a component or in-

terest area is a bit more complex. Generally speaking, the procedure is as follows. All of the items that logically appear to measure a given area, mechanical, for example, are identified, and the test papers of a trial group are scored on these items only. Two subgroups are then chosen, one composed of perhaps the 25 per cent making the highest total scores and another composed of a similar percentage making the lowest scores. These are then known as the *criterion groups,* and item counts are made for each group. The proportion of the "high" group and the proportion of the "low" group responding in a given way on a specific question are compared. To the extent the proportions are the same, the item is dropped from that component; and to the extent they are different, it is retained. This process which is said to employ the criterion of internal consistency is repeated with each component until each item is properly assigned to one or more of the areas.

OCCUPATIONAL GROUP DIFFERENCES

While Sarbin and Anderson[1] have presented some evidence regarding the relationship between expressed job dissatisfaction and measured interests, scales have been most extensively validated by means of occupational groups. Strong has shown significant differences in the scores obtained by persons in various occupations. Doctors, for example, can be expected to score higher on the doctor key than on any other key; as implied earlier, this is inherent in the construction of the scale. Kuder, using nine components on the *Kuder Preference Record,* has demonstrated significant differences in the average patterns of various occupational groups. Figure 7-1, for example, is the group profile for twenty-seven employed chemists. Note that in terms of expressed interest in scientific activities, the group average in the upper 10 per cent of the general population and that they average in the lowest 10 per cent in their interests in clerical activities. Figure 7-2 is the group profile for twenty-

[1] SARBIN, THEODORE R., AND ANDERSON, HEDWIN C. A preliminary study of the relation of measured interest patterns and occupational dissatisfaction. *Educ. Psych. Meas.,* 1942, 2, 23–26.

seven male accountants. Note that their expressed interests are quite high in computational and clerical kinds of activities and

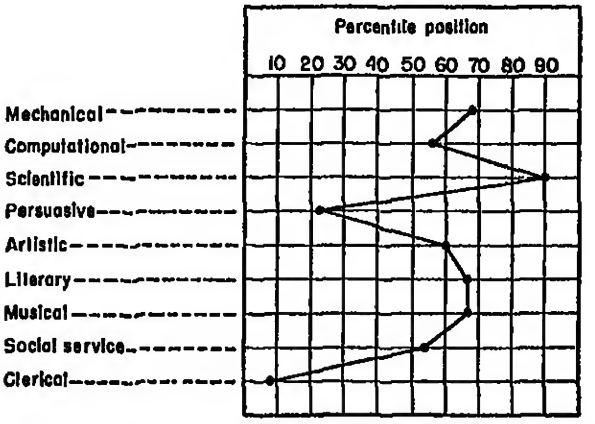


FIG. 7-1.—Group profile based upon mean performance of twenty-seven chemists on *Kuder Preference Record.* (*From Kuder.*)

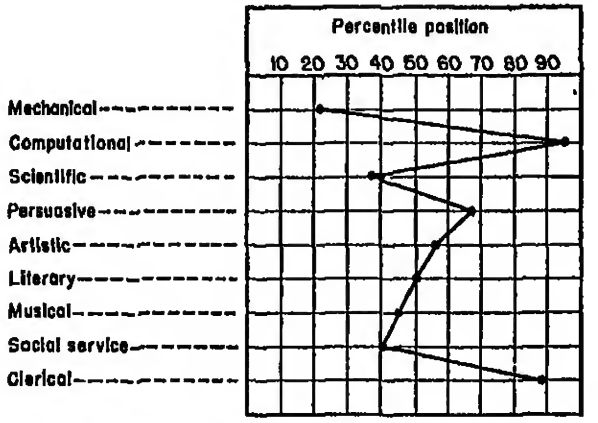


FIG. 7-2.—Group profile based upon mean performance of twenty-seven male accountants on *Kuder Preference Record.* (*From Kuder.*)

that when compared with the general population, they are lowest in their mechanical interests. These two patterns prepared from data selected from Kuder's manual[1] are illustrative of a large

[1] *Intermediate manual for the Kuder preference record.* Chicago: Science Research Associates, 1944.

number that he presents. Generally speaking these group profiles correspond to expected patterns arrived at by logical analysis of the occupation.

ABILITY GROUP PATTERNS

Method Criticized.—The occupational group method for validating interest tests has received some criticism just as it has in connection with mental ability and temperament tests. What evidence is there, say the critics, that one is worse off in

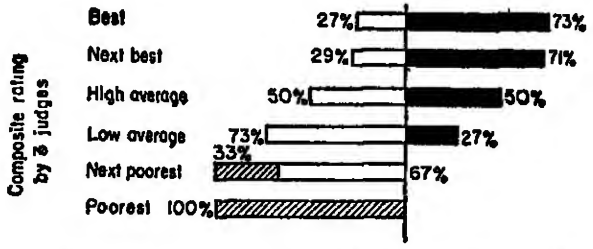


FIG. 7-3.—Proportion of thirty-six advertising men who received A (solid bars), B (open bars), and C (shaded bars) grades on Strong's *Vocational Interest Blank*[1] when grouped according to ratings.

an occupation just because his interests do not coincide with those of the typical employee in that occupation? Strong has answered these criticisms very effectively with considerable research, only part of which will be reported here.

One study[1] involved thirty-six advertising men who were rated as to their occupational competence by three independent judges and their ratings were combined. They were scored on the "advertising man" key of Strong's *Vocational Interest Blank for Men* and given grades of A, B, or C. Figure 7-3 shows the percentage of each rating group that received each test grade. Note that C grades were received only by the poorest and next poorest group and that 100 per cent of the poorest group received C grades. Neither were there any A's in these two bottom groups. The consistent increase in number of A's with succes-

[1] STRONG, EDWARD K., JR. *Vocational interests of men and women.* Stanford University, Calif.: Stanford University Press, 1943, p. 501.

sively higher rated groups is apparent. Strong [1] also has compared the scores of 288 insurance agents, each selling in excess of $150,000 annually, with the scores of twenty-seven individuals who failed as insurance salesmen. Figure 7-4 shows the results. Among the successful group 75 per cent scored A, 24 per cent

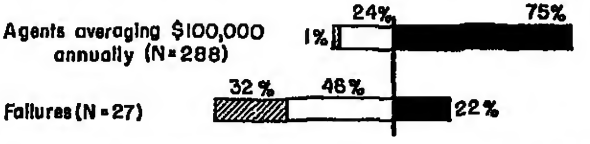


FIG. 7-4.—Proportions of groups of successful and unsuccessful insurance salesmen who received grades of A (solid bars), B (open bars), and C (shaded bars), on Strong's *Vocational Interest Blank*. (*From Strong.*[1])

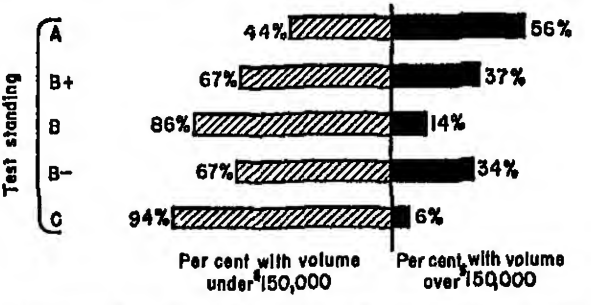


FIG. 7-5.—Proportions of groups of insurance salesmen receiving various grades on Strong's *Vocational Interest Blank* for men who exceeded a sales volume of $150,000. (*From Strong.*[2])

scored B, and only 1 per cent scored C, whereas among the failures 32 per cent scored C, 46 per cent scored B, and only 22 per cent scored A.

In another study involving 211 insurance men, Strong [2] divided the group into two subgroups, those with an annual sales volume below $150,000 and those with a volume in excess of that amount. Figure 7-5 shows the relative proportions of these two subgroups at the various test levels. For example, of those receiving A on the test, 56 per cent were in the high producing group and 44 per cent were in the low producing group. In the same fashion, of those scoring C 94 per cent were low producers

[1] *Ibid.*, p. 480.
[2] *Ibid.*, p. 492.

and only 6 per cent were high producers. The general trend, though somewhat irregular, is nevertheless marked, and there is no question but that superior scores on the insurance key are associated with superior sales performance.

Bills[1] has reported a similar study involving casualty-

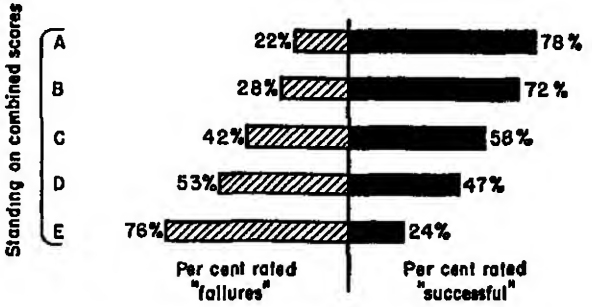


FIG. 7-6.—Proportions of groups of casualty insurance salesmen scoring at various levels on Strong's *Vocational Interest Blank* for men who were rated successful and failures. (*From Bills.*)

insurance salesmen. She investigated 588 salesmen who were rated by their respective managers as failure or successful. She scored each man with the life-insurance-salesman key and the real-estate-salesman key and worked out a combined test score which is represented in Fig. 7-6 by the letters A to E. The figure shows that of those receiving a combined score of A, 78 per cent were successful and 22 per cent were failures whereas of those scoring E only 24 per cent were successful and 76 per cent were failures.

Ryan and Johnson[2] conducted similar studies with a group of men engaged in selling machine accounting methods and another group of accounting machine service or repair men. Figure 7-7 presents their findings for a second group of salesmen after they had previously determined scoring weights with a preliminary group. These salesmen were grouped in terms of job perform-

[1] BILLS, MARION A. Relation of scores in Strong's interest analysis blanks to success in selling casualty insurance. *J. appl. Psychol.*, 1938, 22, 97–104.

[2] RYAN, T. A., AND JOHNSON, BEATRICE R. Interest scores in the selection of salesmen and servicemen: Occupational vs. ability-group scoring keys. *J. appl. Psychol.*, 1942, 26, 543–562.

ance, and the figure shows the percentage of each above and below a critical score. Figure 7-8 shows their findings with the accounting-machine servicemen. The same method of presentation is employed, and the same general trend is prevalent.

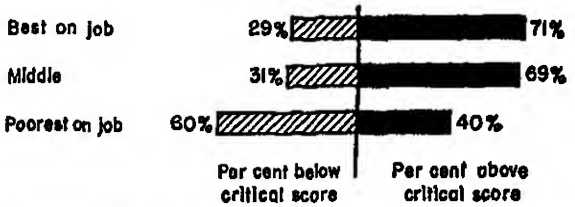


FIG. 7-7.—Proportions of three ability groups of machine accounting methods salesmen who scored above and below a critical score on Strong's *Vocational Interest Blank for Men.* (*From Ryan and Johnson.*)

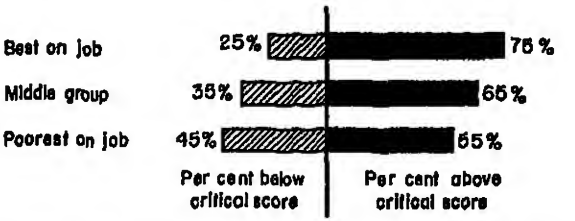


FIG. 7-8.—Proportions of three ability groups of accounting machine servicemen who scored above and below a critical score on Strong's *Vocational Interest Blank for Men.* (*From Ryan and Johnson.*)

LIMITATIONS OF INTEREST TESTS

Guidance Situations.—Interest tests are probably most useful in guidance kinds of situations where the individual taking the test is seeking direction and where all of the influences favor accurate reporting on his part. *There is no question but that individuals can falsify their reports* and that to the extent that they are intelligent and have insight into the nature and demands of jobs they can influence their scores in the desired direction.

However, two facts should be kept in mind. There are many business and industrial situations where placement or direction after the individual has been hired is the real problem. The experience of one company that hires a large number of graduate

engineers is illustrative. Having hired these young men and placed them on the pay roll, the company finds it advantageous to administer the *Kuder Preference Record* to them to assist in directing them into sales work, supervision, research, etc. Here it is to the applicant's best advantage to report accurately.

The second point is one that was also made in connection with temperament testing. Provided the test is always administered in the application situation and provided the test is always validated along the lines outlined in Chap. II, the personnel or employment man cannot go wrong. When this factual approach is used, and if significant differences between acceptable and unacceptable are obtained, how much the applicant falsified is secondary. The criterion is "Does the test work?"

SUMMARY

Interests and preferences are factors in vocational adjustment and job success along with mental ability and temperament. Interest tests or scales are usually one of two types: Either they are scored several times for different occupations, or the scale is scored by interest areas or components and each individual is given a profile. Both have been effectively used. Scales have been validated both by means of occupational groups and by means of ability groups.

CHAPTER VIII

VISUAL SKILL TESTS

Few abilities are as almost universally necessary to the performance of jobs as the ability to see. Although there are jobs that can be done well by blind persons, the vast majority of jobs require that the worker be able to see, even though the visual demands of jobs vary. This chapter deals with vision as it is related to job performance.

THE NATURE OF VISION

Complexity of Vision.—Personnel managers and others responsible for the selection of properly qualified employees have long recognized the importance of vision in connection with job performance as is attested by the frequency with which measures of acuity or keenness of vision have been used. Usually the test has consisted of a letter chart placed about twenty feet away, and more times than not this has been the only test used. The implication is twofold: (1) Acuity or ability to perceive detail is all that there is to seeing, and (2) one's ability to perceive detail at a given distance is an accurate indicator of his ability to perceive detail at other distances. Actually neither of these assumptions is completely true. Figure 2-2 on page 12 is evidence. Employees who did the best job of looping actually had poorer distance acuity than did the poor loopers. Seeing is an extremely complex act. Some professional eye men recognize as many as twenty different visual skills;[1] but whether one accepts the total list or not, research has definitely demonstrated that multiple skills exist and that many of them are relatively unrelated.

Visual Skills and Postures.—The separation of visual skills and visual postures is largely a matter of definition. Generally

[1] SHEPARD, C. F. Visual skills. *Optometric Weekly*, 1944, 34 (51), 1465–1466.

speaking, however, the skills include acuity (or keenness of vision) at various distances, color vision, and stereopsis, a factor in depth perception. Postures included the phorias, lateral and vertical. Measures of vertical phoria indicate the extent to which the two eyes tend to seek a normally level position in contrast to a

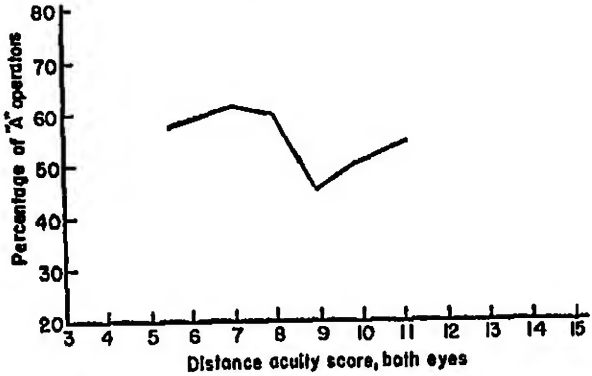


FIG. 8-1.—Relationship between distance acuity and the job performance of 225 radio-tube assemblers. (*From Tiffin and Wirt.*)

position in which one eye tends to assume a position higher or lower than the other. Measures of lateral phoria indicate the extent to which the eyes tend to converge or diverge in the lateral plane.

Relationships between Characteristics.—Geise [1] has demonstrated only a slight relationship between an individual's ability to discriminate detail at various distances. For example, if one knows how well an individual can see at twenty feet, he can estimate how well that person can see at fifteen inches but with a degree of accuracy little better than chance. Geise's findings support previous studies, the results of one of which are presented in Figs. 8-1 and 8-2. Two hundred twenty-five radio-tube assemblers [2] were rated as A or B operators and were also given a standard battery of vision tests. Figure 8-1 shows the

[1] GEISE, W. J. The inter-relationship of visual acuity at different distances. *J. appl. Psychol.*, 1946, 30, 91–106.

[2] TIFFIN, JOSEPH, AND WIRT, S. E. Determining visual standards for industrial jobs by statistical methods. *Trans. Amer. Acad. Ophthal. and Otolar.*, November-December, 1945, 4–25.

percentage of A operators found among those scoring at various levels on a far acuity test (approximately twenty-six feet). Although the curve is somewhat irregular, it can be seen that there are no more A operators among those with the best distance acuity than there are among those operators with poorer distance acuity. If there is a trend, it is in the direction of fewer A operators among those with better distance acuity. Figure 8-2, however, tells quite a different story. The percentage of A

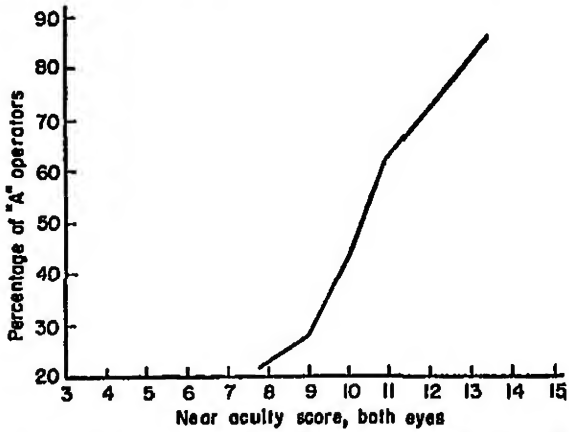


FIG. 8-2.—Relationship between near acuity and job performance of 225 radio-tube assemblers. (*From Tiffin and Wirt.*)

operators increases from about 25 per cent among those who score at the lower end of the near acuity test to approximately 85 per cent among those who scored highest on the same test. These studies along with many others support the conclusion that no single vision test will yield all of the information that is needed for adequate job placement; a battery of vision tests is necessary.

Changes with Age.—Visual skills change with age. Figure 8-3 based upon 7,332 employees in a steel mill shows the percentage of those employees of various ages who passed a far acuity test without the aid of spectacles. Note that the curve is reasonably flat until age forty, at which time it begins to drop systematically. Figure 8-4 indicates a similar trend in unaided near acuity. Color vision similarly deteriorates with age.

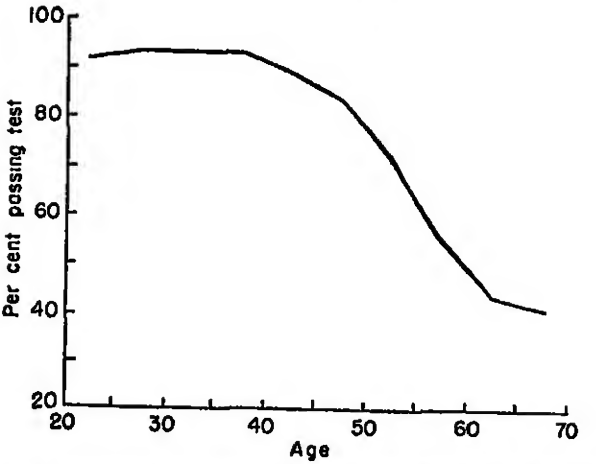


FIG. 8-3.—The percentage of persons at various age levels who were able to pass a certain far acuity test without glasses (N=7,332).

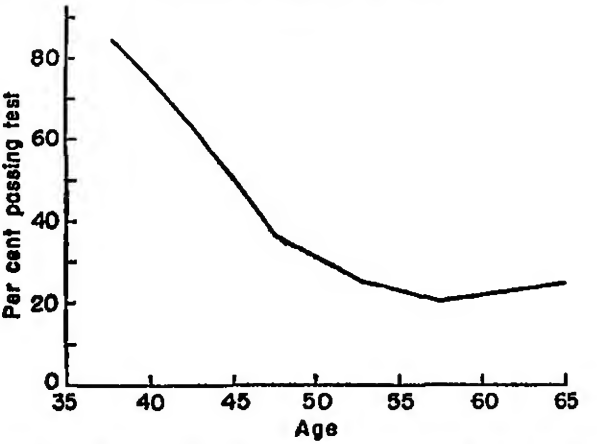


FIG. 8-4.—The percentage of persons at various age levels who were able to pass a certain near acuity test without glasses (N=2,193).

Figure 8-5 from Tiffin and Kuhn [1] shows a decrease of from approximately 70 per cent among twenty-five-year-olds to approximately 25 per cent among sixty-five-year-olds in the percentage of employees passing a red-green color discrimination

[1] TIFFIN, JOSEPH, AND KUHN, HEDWIG S. Color discrimination in industry. *Arch. Ophthal.*, 1942, 851–859.

test. Not all visual skills, however, decline with age in this same systematic fashion. Figure 8-6 shows the changes that take place in stereopsis with age changes. The percentage who could pass a given test increased from about 75 per cent at age twenty-

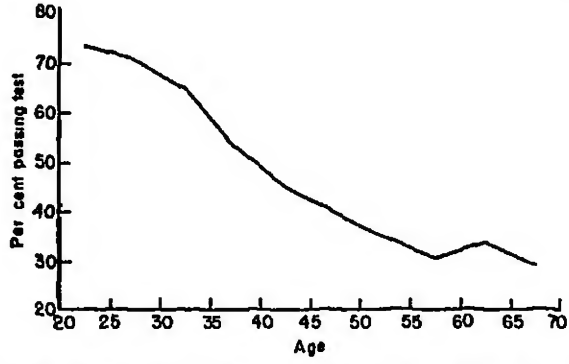


FIG. 8-5.—The percentage of persons at various age levels who were able to pass a red-green color discrimination test (N=7,141). (*From Tiffin and Kuhn.*)

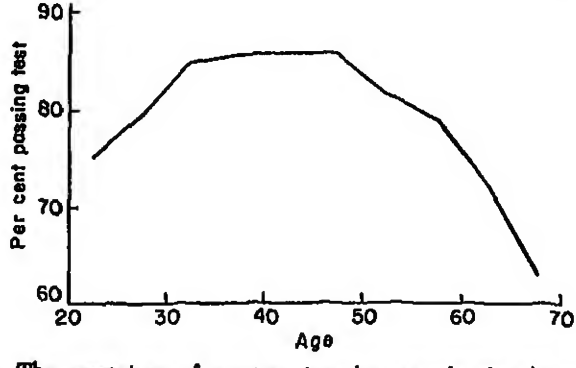


FIG. 8-6.—The percentage of persons at various age levels who were able to pass a certain depth-perception test (N=8,412).

two to about 85 per cent at age thirty-two, remained reasonably constant to about age forty-eight, and then declined. At age sixty-eight, less than 65 per cent of these employees could pass the test. It can readily be seen that to the extent visual skills affected by age are related to success on a particular job, these age shifts are extremely important. Fortunately, vision may be aided through professional eye care. Although there are changes

and deterioration in visual skills with age, adequate professional eye care can to a considerable degree keep visual performance somewhere nearly constant. To the extent that job demands are considered in the giving of professional attention, improvement on the job can result.

Changes through Eye Care.—Kephart [1] has shown how improved vision through eye care may result in improvement on the job. In a study of thirty-one pairers in a hosiery mill who were on the job for a year during which the study extended, seventeen received professional eye care through regular community channels and fourteen did not visit an eye doctor during

FIG. 8-7.—The average cents per hour increase in earnings of seventeen employees who received professional eye care and of fourteen who did not. (*From Kephart.*)

the year. Since employees on this job are paid on a piecework basis, a comparison of earnings accurately reflects shifts in production. The average earnings of all thirty-one employees increased during the year period; but as is indicated in Fig. 8-7 from Kephart's data, those who received professional eye care were averaging 15 cents per hour more at the end of the year as compared with an average increase of only 5.3 cents for those who did not seek eye care in the course of the year.

MEASUREMENT OF VISUAL CHARACTERISTICS

Clinical and Nonclinical Tests.—In basic philosophy, differences exist between clinical eye tests and nonclinical visual performance tests primarily in regard to test construction, test administration, and test application. Optometrists and ophthalmologists use clinical tests to help in diagnosing physiological, pathological, and functional conditions and as aids in determining proper corrective procedures and prescriptions. Such tests usually are flexible to meet the various approaches of the doctor

[1] KEPHART, NEWELL C. An analysis of eye care and industrial efficiency. *Trans. Amer. Acad. Ophthal. and Otolar.*, March-April, 1946, 1–5.

and his clinicians in order to reach the needs of each patient. On the other hand, nonclinical visual performance tests, constructed on a sound psychological basis, should be administered according to an unvarying standard procedure. For this reason, the demands of administering visual skill tests according to standard practice are repetitive and often tedious. Such procedure does not permit opinionated expression of the administrator or the interpretation of any test scores. It follows that thorough training in nonclinical test administration according to standard practice is essential. Only a few, if any, optometrists or ophthalmologists have felt any inclination to do nonclinical testing themselves, the clinical procedures appealing to them because of their temperament and training. Because of standard practice, trained laymen well adapted for this job make successful operators. The training of the layman for the nonclinical test administration ensures professional eye men that diagnosis, prescription, and treatment are the absolute functions of the optometrists or ophthalmologists and can neither effectively nor legally be performed by other.

The Snellen Chart.—The Snellen chart with its characteristic large E is well known to everyone. It is employed by the ophthalmologist and the optometrist as an effective instrument for diagnosis and is widely used as a nonclinical skill test for purposes of recording visual acuity at twenty feet. Although often used as an employment test, its use is limited because it usually measures acuity at one distance only, it measures ability to read letters which places illiterates and literates on different bases, and its results are subject to external factors such as quality and quantity of light. A battery of tests sufficiently varied to inventory those visual skills most generally important in job success is needed.

Non-clinical Tests.—Three different vision testing instruments, nonclinical in nature and designed for classification purposes, are at present in use in industry. These are all binocular testing instruments[1] and are called the *Telebinocular*, the *Sight*

[1] *Methods of testing and protecting eyesight in industry.* Industrial Health Series No. 4. New York: Metropolitan Life Insurance Co., 1945.

Screener, and the *Ortho-Rater.* Each of these instruments provides for presenting to the employee a series of "targets" in order to measure his visual performance. The targets employed and the visual skills measured vary somewhat from instrument to instrument. The manufacturers of the instruments claim administration times for their batteries that range from three to five or six minutes. Other differences pertain to methods of installation, operation, and administration rather than to instrumentation.

THE VALIDATION OF VISION TESTS

Testing the Test.—As outlined in Chap. II, no test should be used for the selection and allocation of employees or applicants until that test has been validated or tried out with employees on that particular job. Vision tests are no exception to this rule. Figure 2-2 cited earlier demonstrates not only what can happen but what has happened when wrong vision tests were used to establish job standards. The establishment of visual standards on any basis other than statistical not only is apt to limit the available personnel without improving those hired but may actually result in the placement of inferior employees on the job. Published results which demonstrate significant relationships between measured visual skills and job success in industry are almost entirely based upon research conducted with the *Ortho-Rater,*[1] and it is the only one of the instruments upon which there is published reliability data on industrial workers. Since all of the following statements show relationships between job success and scores made on the *Ortho-Rater,* a more complete description of the instrument is desirable. The *Ortho-Rater* is illustrated in Fig. 8-8 and includes twelve distinct visual skill tests, each of which has been proved to measure a function that is important for successful job performance in several types of occupations. Seven of the twelve are given at the optical equivalent of twenty-six feet, and the remaining five at thirteen inches. The tests are

[1] Distributed by the Industrial Vision Department, Bausch and Lomb Optical Co., Rochester, N.Y.

At twenty-six feet:
- Vertical phoria
- Lateral phoria
- Acuity, both eyes
- Acuity, right eye
- Acuity, left eye
- Depth perception
- Color vision

At thirteen inches:
- Acuity, both eyes
- Acuity, right eye
- Acuity, left eye
- Lateral phoria
- Vertical phoria

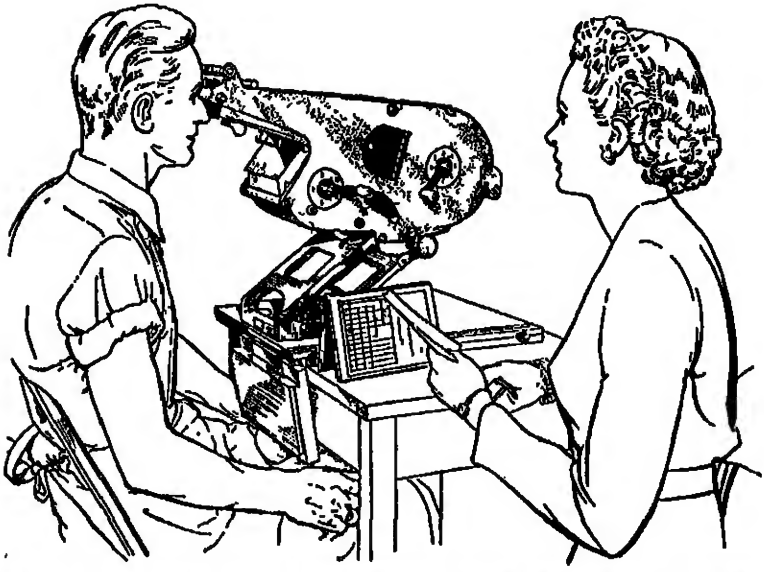


FIG. 8-8.—The *Ortho-Rater*, a nondiagnostic testing instrument. (*Courtesy of Bausch and Lomb Optical Co.*)

The instrument is nonclinical and is usually operated by laymen who are trained in the use of a standard testing procedure.

The Individual Profile.—As each employee or applicant is tested, a visual profile similar to the one in Fig. 8-9 is prepared. When job standards are established, it is quite easy to determine whether or not a given individual meets those standards.

Distance Acuity.—Coleman [1] has reported a study in which the importance of the distance acuity test for both eyes in con-

[1] COLEMAN, J. H. Vision tests for better utilization of manpower. *Fact. Mgmt. and Maint.*, July, 1944.

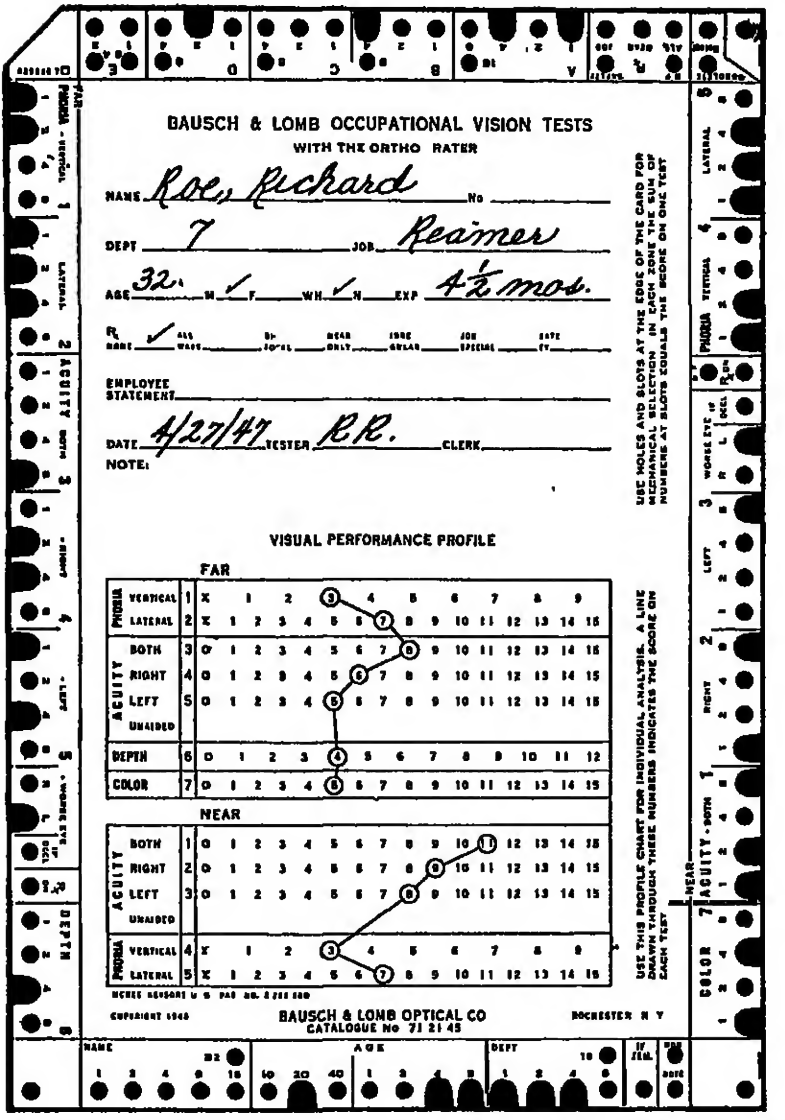


FIG. 8-9.—Sample individual record card showing visual profile of one employee.

nection with ninety-seven milling-machine operators is demonstrated. The operators were rated as superior, good, fair, or poor by the supervisors, and Fig. 8-10 based on Coleman's report indicates the proportion of each rated group that attained a test score of 9 or more. Whereas only half, or 50 per cent, of the poor rated group met this standard, 84 per cent of the superior group attained this level. Further analysis of his results indicates that although 49 per cent of the whole group were rated

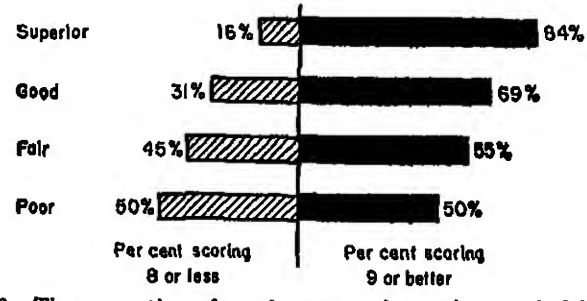


FIG. 8-10.—The proportion of employees rated superior, good, fair, and poor who attained a distance acuity score of 9 or better. (*From Coleman.*)

good or superior, if only those with far acuity scores of 9 or better are considered, 58 per cent received one of these better ratings.

Near Acuity.—Many industrial jobs require superior near point acuity. Tiffin and Wirt[1] have reported a study showing the relationship between near acuity scores and the earnings of seventy-two electric solderers. These solderers had near acuity scores ranging from 7 to 15, and their piecework earnings averaged approximately 80 cents per hour. Figure 8-11 shows the percentage of those at each score level who earned 80 cents or more per hour for the period studied. The figure shows that none of those scoring 7 or 8 received 80 cents per hour and that the higher the vision score the greater the proportion of high earners. About 65 per cent of those scoring 12 on the test were among the high earners.

Worse Eye.—Some jobs may be done well by persons with

[1] TIFFIN, JOSEPH, AND WIRT, S. E. Determining visual standards for industrial jobs by statistical methods. *Trans. Amer. Acad. Ophthal. and Otolar.*, November-December, 1945, 4-25.

one eye or by persons who have low acuity in one eye. Others, however, require a minimum acuity in each eye. Some studies utilizing *Ortho-Rater* data have included the checking of "worse eye" acuity scores against measures of job success. Tiffin [1] has reported such a study, and Figs. 8-12 and 8-13 are reproduced

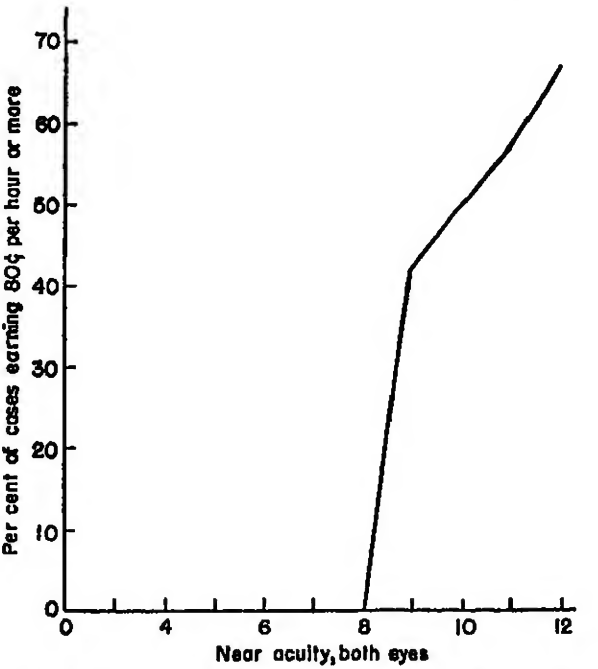


FIG. 8-11.—The relationship between near acuity scores and earnings of electric solderers. (*From Tiffin and Wirt.*)

from his report dealing with a group of piston ring inspectors. The operators on the job were rated as A, B, C, or D on the basis of production data. Fifty-three per cent of the whole group were A or B operators, and Fig. 8-12 shows the proportion at various levels of distance acuity (worse eye). Tiffin points out that there is a relatively slight change in the proportion of A and B operators in the 1 to 5 bracket but that for higher scores there is marked percentage increase and that approximately 80

[1] TIFFIN, JOSEPH. The use of visual data as an aid to increase production and efficiency. *Trans. Amer. Acad. Ophthal. and Otolar.*, January-February, 1944.

per cent of those scoring 11 were superior. Figure 8-13 shows the corresponding pattern for near acuity (worse eye). Whereas essentially the same pattern is found, the obtained percentage is

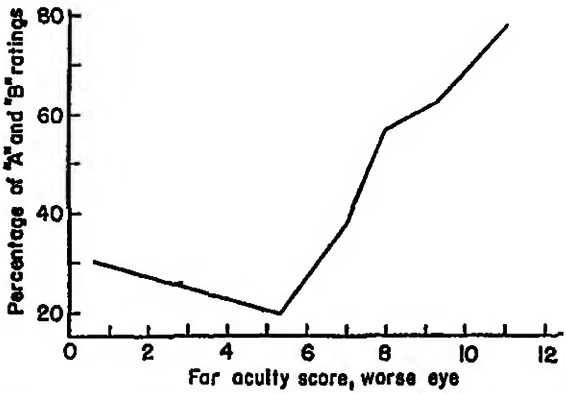


FIG. 8-12.—The relationship between rated job performance on piston-ring inspection and distance acuity test score for worse eye. (*From Tiffin.*)

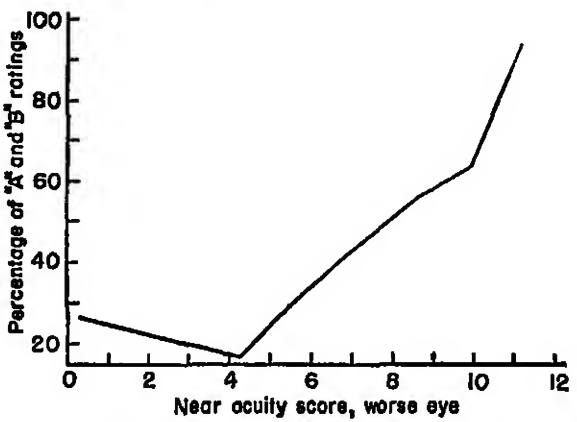


FIG. 8-13.—The relationship between rated job performance on piston-ring inspection and near acuity test score for worse eye. (*From Tiffin.*)

even greater among the high scorers, 96 per cent of those scoring 11 having been considered A or B inspectors.

Phoria.—The phoria tests, sometimes referred to as the posture tests, measure the position that the eyes tend to assume in a condition of physiologic rest. In the *Ortho-Rater* both the

lateral phoria test[1] and the vertical phoria test are constructed so that orthophoria (the tendency for neither eye to deviate) is indicated by the middle scores. Consequently, with most jobs

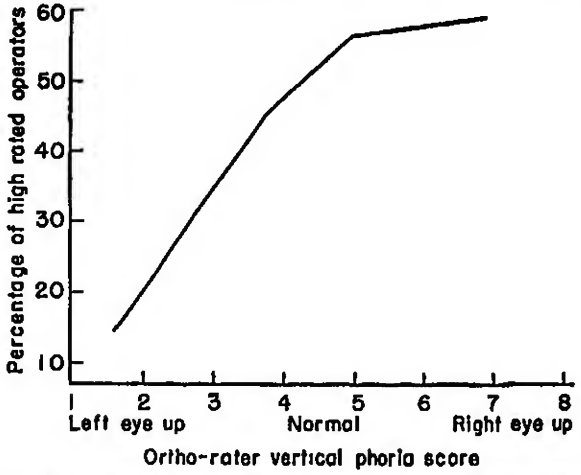


FIG. 8-14.—The relationship between near vertical phoria and rated success of ninety-five milling-machine operators. (*From Tiffin and Wirt.*)

where muscular balance is important, middle scores constitute the acceptable range and extremely high or low scores constitute the unacceptable range. Figure 8-14 taken from Tiffin and Wirt,[2] however, pertains to a job that is unique in this respect. The figure based upon ninety-five milling-machine operators shows a greater proportion of high rated operators among those whose eyes are well balanced vertically or whose right eye tends to tilt upward slightly. A small degree of right eye upward deviation can be tolerated, but no tolerance for the left eye is indicated. This fact, while probably associated with the head position assumed by many operators, is still another example of the need for the statistical determination of visual standards for each job.

Color.—Color vision is known to be important in many jobs

[1] WIRT, S. EDGAR. Studies in industrial vision. I. The validity of lateral phoria measurements in the Ortho-Rater. *J. appl. Psychol.*, 1943, 27, 217–232.

[2] TIFFIN, JOSEPH, AND WIRT, S. EDGAR. The importance of visual skills for adequate job performance in industry. *J. consult. Psychol.*, 1944, 8, 80–89.

where obviously the operator must make color discriminations as is the case of certain printers who work with colored inks. Less obvious is the association of poor job performance with color-discrimination deficiency in many jobs where no logical job analysis would reveal the fact. Stump[1] showed a relationship between accident information and *Ortho-Rater* color test scores. Accident records of a group of employees were examined and

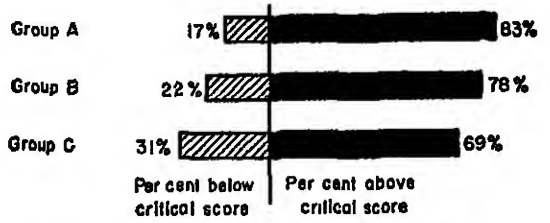


FIG. 8-15.—Proportions of an *accident-free group* (A), a *high-frequency* accident group (B), and a serious injury group (C) who attained and did not attain a critical score on the color-vision test. (*From Stump.*)

three classifications were set up: Group A, accident free; Group B, high frequency; and Group C, serious injury. Figure 8-15 based upon Stump's data shows the proportion of each group attaining and not attaining a critical score of 4 on the *Ortho-Rater* color tests. Eighty-three per cent of the accident-free group passed the standard, but only 69 per cent of the serious-injury group attained the standard. These differences, though not large, are significant and have been found in many jobs that do not appear to demand color vision.

THE VISUAL PROFILE

Job Variability.—It has been falsely and expensively assumed that all industrial jobs have required one "good vision," but it is now known that jobs differ greatly in their specific visual requirements. When the validation techniques presented in Chap. II are applied to each of the twelve *Ortho-Rater* tests for a number of jobs, marked differences from job to job are found. Far

[1] STUMP, N. FRANK. Spotting accident-prone workers by vision tests. *Pact. Mgmt. and Maint.*, June, 1945.

acuity, both eyes, is important on one job and not on another, whereas near lateral phoria is important on the first job but not on the second. The result is considerable variation from job to job in the statistically determined visual patterns which can best be shown as job profiles.

Job Profiles.—Just as visual profiles for individuals are prepared, visual profiles for jobs[1] can also be prepared. As each of

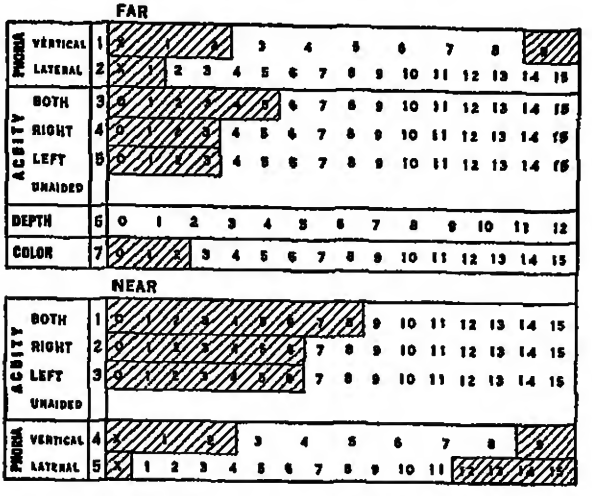


FIG. 8-16.—Job profile for electric solderers. Scores in shaded areas are unacceptable. (*From Stump.*)

the twelve tests is subjected to the validation procedure in connection with a specific job, those tests which are significant for that job are identified and appropriate "cutoffs" or critical scores are established. By shading the unacceptable areas and leaving the acceptable zones or ranges white, it is a simple matter to match any given individual's profile against the profile for that job and to determine to what extent he possesses the visual characteristics of the more successful employees on that job. In actual practice the job profiles are prepared on transparent templates so that individual profiles may be placed underneath and

[1] WIER, S. E. Statistical laboratory for vision tests at Purdue University. *J. appl. Psychol.*, 1946, 30, 354-358.

a quick determination made. Figures 8-16 and 8-17 show the statistically determined job profiles for an electric soldering operation and a milling-machine operation, respectively. A test-by-test comparison indicates that the individual whose profile appears on page 105 as Fig. 8-9 meets the standards for electric soldering presented in Fig. 8-16; in every case his scores fall in

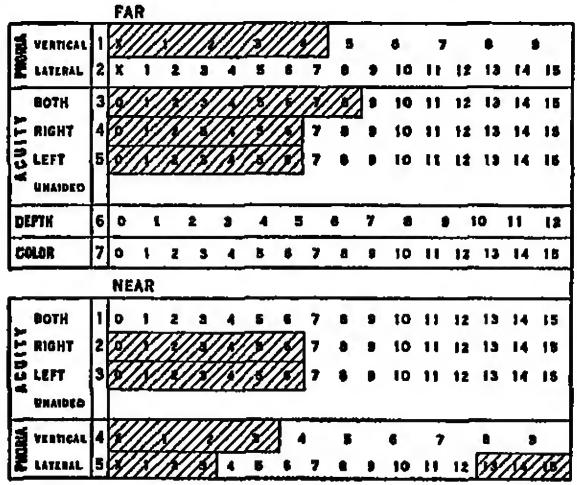


FIG. 8-17.—Job profile for milling-machine operators. Scores in shaded areas are unacceptable. (*From Coleman.*)

the white area, and they never fall in the shaded area. Quite a different picture appears when this same individual profile is matched against the job profile for milling in Fig. 8-17. Note that this individual fails to meet the standards on five different tests: far vertical phoria, far acuity both, far acuity right, far acuity left, and near vertical phoria. Other things being equal, this individual as an applicant is a poor risk on the milling job because he does not possess the visual characteristics of the better operators. As a present employee, there is a good chance that his job performance can be improved by professional eye care. One point is quite clear, however: The visual demands of jobs differ to such an extent that almost any individual's visual profile will fit somewhere.

Electric Solderers.—The job profile for electric solderers

shown in Fig. 8-16 is reproduced from Stump's report.[1] These solderers were paid on a piecework plan, and Fig. 8-18 shows the percentage at each income level who met the standards. Of those who earned 60 cents per hour or less (even though they were paid a minimum wage) 40 per cent met the standards shown in Fig. 8-16, whereas of those earning 90 cents per hour or more

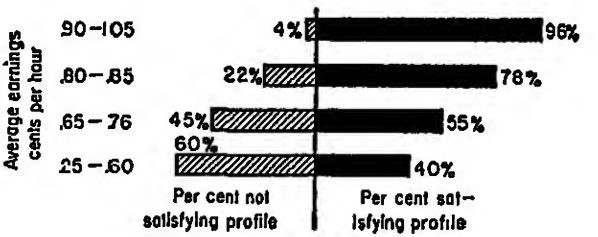


FIG. 8-18.—Percentage of electric solderers in each earning bracket who passed and failed visual standards presented in Fig 8-16. (*From Stump.*)

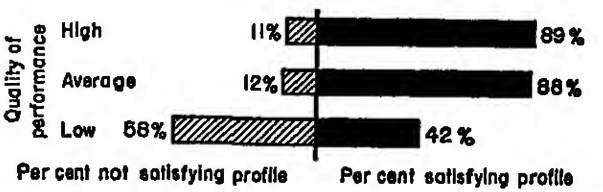


FIG. 8-19.—Percentage of electric solderers in three quality performance brackets who passed and failed visual standards presented in Fig. 8-16. (*From Stump*)

96 per cent met the standards. The figure also shows a consistent pattern for the other pay levels. Stump further reports that all those employees who failed to meet the job profile pattern averaged 60 cents per hour whereas those who did meet the standards averaged over 80 cents per hour; in other words, those who meet the standards earn on the average about 34 per cent more than do those who do not meet the standards.

Supervisory ratings on quality of work performed by the solderers were available. Figure 8-19, also from Stump's study, shows that 89 and 88 per cent respectively of those rated high and average on quality met the standards but only 42 per cent

[1] STUMP, N. FRANK. Vision tests predict worker capability. *Fact. Mgmt. and Maint.*, 1946, 104, 121–124.

of the low-quality employees met the standard. Stump further reports that for a thirteen-week period studied, those failing the standards averaged 28.6 half days of absence as compared with 17.1 half days for those who met the standards.

Milling-machine Operators.—The job profile for milling-machine operators shown in Fig. 8-17 is from a study reported

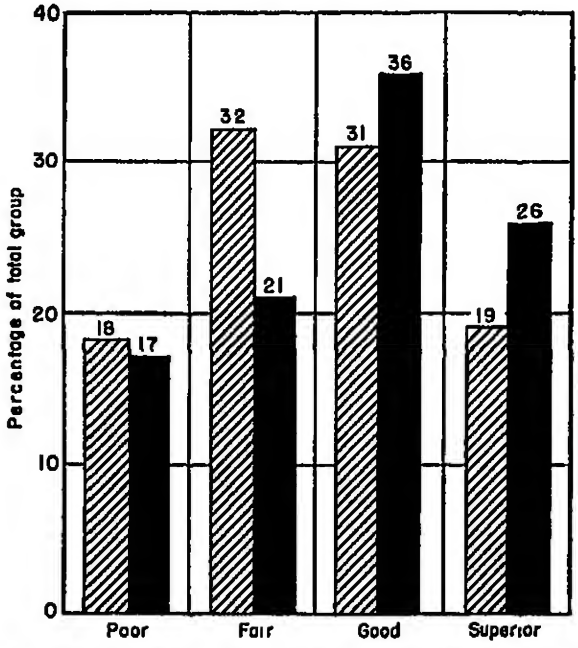


FIG. 8-20.—Percentage of all milling-machine operators who were rated in four categories (solid bars) vs. percentage of those who met visual standards (Fig. 8-17) who were so rated. (*From Coleman.*)

by Coleman.[1] These operators were rated as superior, good, fair, and poor, and the shaded bars in Fig. 8-20 shows that the percentage distribution among all of the operators was 19, 31, 32, and 18 per cent, respectively. However, when only those meeting the standards are considered, percentages of 26, 36, 21, and 17 per cent represented by the black bars were obtained. A significantly higher number fall in the good and superior brackets

[1] COLEMAN, J. H. Vision tests for better utilization of manpower. *Facts. Mgmt. and Maint.*, July, 1944.

and a significantly smaller percentage fall in the fair and poor bracket when only those meeting the standards are considered.

Piston-ring Inspectors.—Figure 8-21 shows the job profile that was statistically established for a group of piston-ring inspectors.[1] Although only four of the tests "came through" in this particular study, Fig. 8-22 shows an increase in the proportion

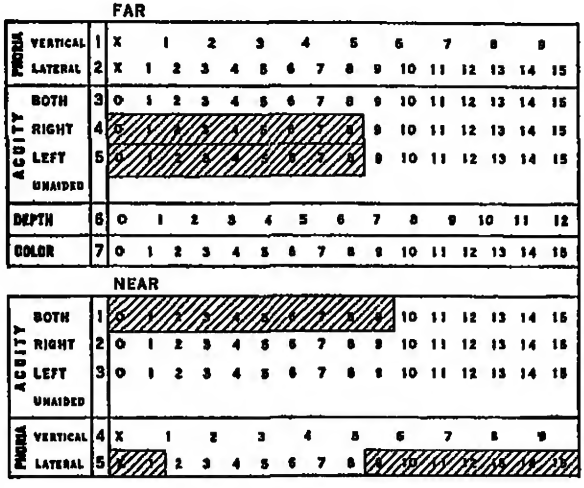


FIG. 8-21.—Job profile for piston-ring inspectors. Scores in shaded areas are unacceptable. (*From Tiffin.*)

of A and B operators and a decrease in the proportion of C and D operators when only those who meet the standards are considered.

Drill-press Operators.—Figure 8-23 shows the standards[2] of drill-press operators using jigs. Here only two tests, lateral phoria near and lateral phoria far, were identified as being important. Figure 8-24 shows in A, B, C, and D operators the proportional shift that occurs when only those who meet the established visual standards are considered. Note the increase in A and B operators and the decrease in C and D operators.

The Validation Approach.—For further emphasis it seems desirable to point out again that the visual standards for the four

[1] TIFFIN, JOSEPH. Vision and industrial production. *Illum. Engin.*, 1945, 40, 239–257.

[2] *Ibid.*

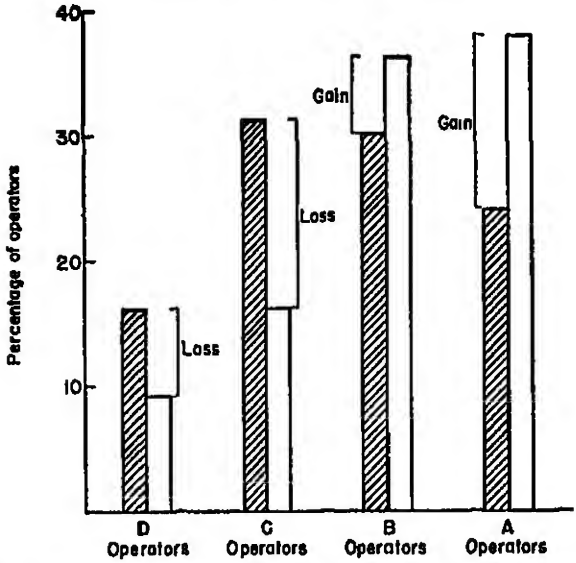


FIG. 8-22.—Percentage of all piston-ring inspectors who were rated in four categories vs. percentage of those who met visual standards (Fig. 8-21) who were so rated. (*From Tiffin.*)

FAR

PHORIA	VERTICAL	1	X	1	2	3	4	5	6	7	8	9				
	LATERAL	2	X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ACUITY	BOTH	3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	RIGHT	4	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	LEFT	5	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	UNAIDED																	
DEPTH		6	0	1	2	3	4	5	6	7	8	9	10	11	12			
COLOR		7	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

NEAR

ACUITY	BOTH	1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	RIGHT	2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	LEFT	3	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	UNAIDED																	
PHORIA	VERTICAL	4	X	1	2	3	4	5	6	7	8	9						
	LATERAL	5	X	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15

FIG. 8-23.—Job profile for drill-press operators. Scores in shaded areas are unacceptable. (*From Tiffin.*)

job classifications discussed here are not arbitrary estimates but have been systematically arrived at by means of the general validation procedure discussed in Chap. II. In every instance, good and poor employees have been identified by one or more of the means listed in Chap. III, and the performance of each group on

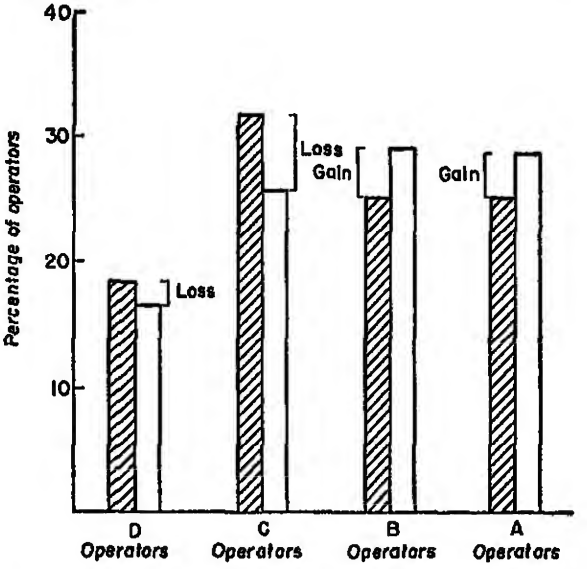


FIG. 8-24.—Percentages of all drill-press operators who were rated in four categories vs. percentage of those who met visual standards (Fig. 8-22) who were so rated. (*From Tiffin.*)

each of the twelve tests was compared. Only those tests which showed differences between these two groups have been included in the standards. The variability of the standards from job to job is quite apparent. Already standards[1] for between 700 and 800 job classifications have been established.

Visual Job Families.—A more recent development in the establishment of standards has been extremely useful in those departments or plants where there is a small number of employees in any one job classification, making the statistical establishment

[1] These have been established by the Occupational Analysis Laboratory, Division of Applied Psychology, Purdue University. An unknown number have been established by industries themselves.

of standards for these individual classifications practically impossible. Techniques beyond the scope of this discussion have been developed whereby different jobs in the same department with similar visual demands can be grouped together into visual job families and standards established for the family rather than for the individual classification. This grouping procedure which is a statistical process sometimes classified jobs in a particular situation into two visual job families with quite similar standards, one of which is more rigorous than the other.

SAFETY AND VISION

Accident Proneness.—The accident literature presents much evidence in support of the statements that a small proportion of individuals have a high proportion of the accidents, that this small group has more accidents than can be accounted for on the basis of chance alone, and that the more accidents a given individual has had during a previous period the greater his probability of repeating during the next period. These persons who have more than their share of accidents have come to be called "accident prone," and accident proneness simply describes the fact that an individual has more accidents.

Vision as an Accident Cause.—Although all of the characteristics of accident-prone employees are not known, sufficient progress has been made that Stump [1] has estimated on the basis of evidence that about one-fourth of industrial injuries could be eliminated if proper standards of visual performance were established and adhered to. Stump,[2] in another report, describes the establishment of *Ortho-Rater* visual standards using degree of accident proneness as a criterion. By means of visual standards so established, it was possible to classify employees on the basis of vision scores alone.

Wirt and Leedke [3] have reported significant relationships be-

[1] STUMP, N. FRANK. Visual functions and safety. *National Safety News*, June, 1944.

[2] *Ibid.*

[3] WIRT, S. EDGAR, AND LEEDKE, HAZEL N. Skillful eyes prevent accidents. *Annual News Letter*, Industrial Nursing Section, National Safety Council, November, 1945.

tween vision scores and the accident records of paper-machine operators and tradesmen. In the case of the paper-machine operators, fifty-two men who had a record of one or more serious accidents and another fifty-two with no serious accidents were matched on the basis of age and experience on the job. The *Ortho-Rater* scores of these men were matched against the job

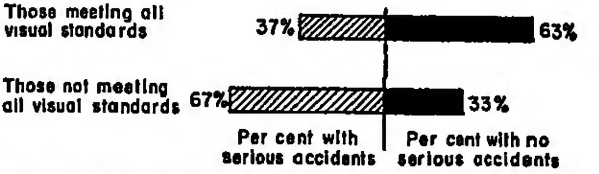


FIG. 8-25.—Percentage of paper-machine operators meeting visual standards and not meeting visual standards who had serious accidents. (*From Wirt and Leedke.*)

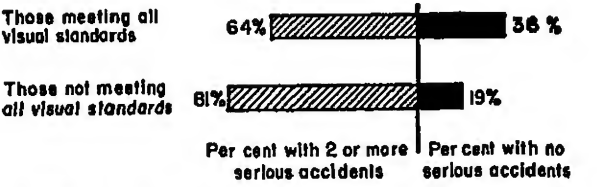


FIG. 8-26.—Percentage of millwrights meeting visual standards and not meeting visual standards who had two or more serious accidents. (*From Wirt and Leedke.*)

profile previously established in accordance with the procedure already outlined. These 104 employees were then classified as having met or not having met all visual standards for that job. Figure 8-25 shows that whereas 67 per cent of those not meeting all visual standards had one or more accidents, only 37 per cent of those who met the standards had accidents. Similar results were obtained in the study involving a group of ninety-four millwrights and other tradesmen. Figure 8-26 shows that whereas 81 per cent of those who failed to meet all visual standards had two or more accidents, among those who did meet every standard, 64 per cent had two or more accidents.

Safety as a Measure of Job Success.—As pointed out in Chap. III, measures of job success that are used in test-validation

procedures frequently reflect the nature of the personnel or production problems which are associated with a given job or a particular plant. Similarly the accident problem is one that is of considerably greater importance on certain jobs and in certain industries than in others. The importance of accident records should not be overlooked when ways and means of classifying employees for purposes of test validation are being sought.

Selection and Placement.—Since this book deals with psychological tests for purposes of personnel selection and placement, the emphasis in this chapter has been primarily on selection and placement. However, no area of psychological testing is any more far-reaching in its implication for all-around personnel improvement than is vision testing. It seems important to discuss related phases of a total industrial vision program.

Professional vs. Nonprofessional Services.—The relationship between the lay tester and the professions has been clearly seen by Kuhn.[1] A pioneer in the medical profession in the use of job analysis in the correction of visual defects, she has demonstrated professional participation on a high level. There is no conflict between professional activity of this sort and those functions which can adequately and effectively be performed by the trained lay test administrator. Those types of testing which are clinical in nature are clearly among the activities that must be performed by the ophthalmologist and optometrist.

Referrals.—The lay tester cannot diagnose or prescribe. Therefore it is desirable that he refer to ophthalmologists and optometrists persons whose visual skills do not fall within the acceptable limits that have been established by statistical means for that job. Whether the plant has a full-time medical director, employs an optometrist or an ophthalmologist as a consultant, or leaves the employee to seek in the community the service that he needs has little bearing on the fundamental principles of operation. Most important of all is the fact that the professional man be made thoroughly familiar with the testing procedure, the benefits of lay testing, and the meaning of job standards as statistically established.

[1] KUHN, HEDWIG S. *Industrial ophthalmology*. St. Louis: Mosby, 1944.

Safety Eyewear.—Whitney [1] has reported that in one plant injuries to the eyes constituted 12 per cent of all of the injuries for a six months' period. The importance of eye protection through safety eyewear is well known to safety engineers. Logically an eyewear program should be tied in with the total program. Far too often safety goggles are more or less forced on the employee without any regard for the effect that they might have on his visual performance. Vision and safety are related; therefore attention to eye protection without consideration of eye performance is unsound. A safe employee may even be made an unsafe employee through inadvertent interference with his visual performance. Feinberg and Sewell [2] have described one five-point program as follows:

1. Preliminary testing of twelve visual skills of each employee to determine those in need of further examination.
2. Provision of a complete eye examination for those requiring it as indicated by above tests.
3. Provision of safety glasses, either plain or prescription, correctly fitted by a company optician, without cost to the employee, with the provision that he would be responsible for the glasses and that they remain the property of the company.
4. Extension of preliminary visual testing to the employment office for preplacement evaluation of applicants.
5. Furnishing safety glasses to all visitors to the areas covered by the directive.

Eye protection [3] is a scientific matter, and any eyewear program rightly deserves the guidance of a competent professional man.

Job Improvement.—Even though eyes were made before jobs, many jobs have been designed or engineered with little or no thought of the demands that they make on the human being.

[1] WHITNEY, L. HOLLAND. Industrial first aid—eye injuries. *Industr. Med.*, 1945, 14, 337–338

[2] FEINBERG, RICHARD, AND SEWELL, JOHN C. Eyesight through foresight. *National Safety News*, January, 1945.

[3] For a comprehensive discussion see KUHN, *op. cit.*, pp. 173–221.

Many jobs are capable of visual simplification. The discussion of this problem is beyond the scope of this book, but it should be pointed out that the increased emphasis on work simplification as well as the work that is being done in the setting of visual job standards is sufficiently spotlighting this phase of engineering so that more and more thought is being given to the employee and his eyes.

SUMMARY

Vision testing is a highly important phase of the total personnel testing program. There is no one "good vision," and it has been demonstrated that jobs do not all make the same visual demands upon the employee. There is a great need for the application of the same kind of validation procedures to vision tests as have been advocated for other tests. There is a distinction between clinical and nonclinical tests; and whereas the former should be administered only by an ophthalmologist or an optometrist, the latter can be administered to advantage by the layman provided he is adequately trained. There is a real opportunity for coordinating the vision-testing program and those phases of the safety program which encompass vision.

CHAPTER IX

TESTS FOR MECHANICAL AND OTHER MANUAL WORKERS

The group of employees classified as mechanical workers constitutes the largest single category of industrial personnel. Tests intended to predict success on jobs in this area are consequently of extreme importance. The purpose of this chapter is to report significant validity studies in this field and to enumerate and describe certain available tests.

The Measurement of Mechanical Ability.—The terms *mechanical aptitude* or *mechanical ability* are used with such a variety of meanings that they are almost meaningless without some attempt at definition. So-called "mechanical" jobs range throughout the various occupational levels, and the term mechanical aptitude is frequently used to describe the individual who tends to excel, regardless of the occupational level. Although research has not identified the components of mechanical ability, it is quite probable that there are three reasonably distinct components.[1]

Probably the most important component is associated with the capacity or the ability to understand mechanical relationships and mechanical processes of various sorts. Engineers are probably at the very top of this particular component.

Manipulative ability[2] constitutes a second component. It probably embraces the various dexterities and muscle coordinations; and while certain research has demonstrated low interrelationships between measures of certain of these abilities, it

[1] Bennett, George K., and Cruikshank, Ruth M. *A summary of manual and mechanical ability tests.* New York. Psychological Corp., 1942.

[2] Long, W. F., and Lawshe, C H, Jr The effective use of manipulative tests in industry. *Psychol. Bull.*, 1947, 44, 130–148.

may be that there is a generalized manipulative or muscular factor.

Still a third element that may be included in an analysis of mechanical aptitude is involved in the motor abilities of strength, speed of movement, and endurance. But whether future researches support or refute this classification, it is nonetheless useful at the present time for thinking through the use of tests for industrial purposes.

Tests of Mechanical Ability.—Tests of mechanical ability may be classified in a variety of different ways, one of which is in terms of the nature of the testing materials themselves. Tests classified in this way fall into one of two categories: paper-and-pencil tests and apparatus tests. Certain elements can be measured either by apparatus or by paper-and-pencil tests, but each type has unique purposes for which it alone can be used.

Tests of mechanical ability may also be classified in terms of their function or intended function. Bennett and Cruikshank [1] list four groupings which seem to include all paper-and-pencil tests. These groupings are listed below.

Information.—Tests of this sort range all the way from those which measure general background knowledge such as basic tool information to rather detailed tests of specific trade information.

Spatial Relations.—Tests that measure space and form perception have been shown to correlate with ability to perform some kinds of mechanical work. Some authorities contend that tests of this sort are really nonlanguage intelligence tests.

Mechanical Comprehension.—Tests of this character measure more than factual knowledge as such. They tend to tap an individual's understanding of underlying mechanical principles and relationships.

Interest.—Most interest tests such as those discussed in Chap. VII include mechanical areas or components.

Certain of the apparatus tests may also be used to provide measurement in some of the aforementioned areas, particularly spatial relations and mechanical comprehension.

[1] BENNETT AND CRUIKSHANK, *op. cit.*

JOB KNOWLEDGE OR TRADE TESTS

Definition.—When a teacher who has offered instruction in arithmetic wishes to evaluate the student's learning in that field, he administers an arithmetic-achievement test. A job-knowledge or trade test is fundamentally an achievement test. It measures the individual's learned knowledge or skill in a particular occupational area in which he has had either experience or specific training. Although most trade tests are information tests using oral or written questions, some trade tests are manipulative in nature and measure performance skill. Regardless of their nature, however, they are primarily useful in evaluating an individual's present adequacy in the job area.

Oral Trade Questions.—A trade test utilizing oral trade questions is basically no different from a pencil-and-paper trade test. It is sometimes argued that since people in certain occupations seldom engage in reading and writing in connection with their jobs, they are ill at ease and at a disadvantage when asked to take a written test. The validity of this argument has not been demonstrated. However, oral trade questions are useful, and Thompson[1] in 1936 published an oral trade question manual. However, by far the most comprehensive and ambitious work in this area was done by Stead, Shartle, and associates[2] of the United States Employment Service. The Worker Analysis Section of this agency did extensive work in the oral trade question area and made impressive manuals available to the various local U.S.E.S. offices. They have not been made available to private industry and business for obvious reasons. One example, however, is sufficient to demonstrate the usefulness of this approach. In the oral trade test for painters, fifteen questions were finally retained by methods similar to those described in Chap. II. In the validation process, these fifteen questions were asked of a group of expert painters, a group of apprentices and helpers,

[1] THOMPSON, LORIN A., JR. *Interview aids and trade questions for employment offices.* New York: Harper, 1936. 173 pp.

[2] STEAD, WILLIAM H., SHARTLE, CARROLL L., *et al. Occupational counseling techniques.* New York: American Book, 1940.

and a group of related workers. The term *related workers* was used to identify a group of workers who, though not actually engaged in the job in question, were in a favorable position for picking up certain incidental information about the job. Figure 9-1 shows the performance of these three groups on the fifteen questions. The possible score range was divided into three brackets, namely, 0 to 5, 6 to 8, and 9 to 15. The figure indicates that whereas 78 per cent of the expert painters scored in the 9 to

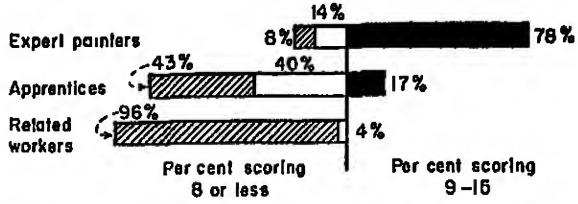


FIG 9-1.—Percentage of three classes of employees scoring 9 to 15 (solid bars), 6 to 8 (open bars), and 0 to 5 (shaded bars) on an oral trade test for painters. (*From Stead and Shartle.*)

15 bracket, only 17 per cent of the apprentice group and none of the related worker group scored in the category. The lowest category, 0 to 5, included 96 per cent of the related workers, 48 per cent of the apprentice group, and only 8 per cent of the expert painter group.

It is probable that oral trade questions will be most useful in the future when they are developed for use in specific industries. It is quite likely that the use of such tests for upgrading purposes has not yet been sufficiently explored.

Written Trade Tests.—As indicated above, trade tests of the paper-and-pencil variety are essentially the same as oral trade questions. Because they can be administered to groups, they are more easily standardized and are more comprehensive.

Trade tests are available for a number of occupations, and reference to Appendix C will reveal sources of supply. Specific mention is made here, however, of the series known as the *Purdue Vocational Tests* [1] and a later series called the *Purdue Personnel*

[1] Published by Science Research Associates, Inc., 228 South Wabash Ave., Chicago 4, Ill.

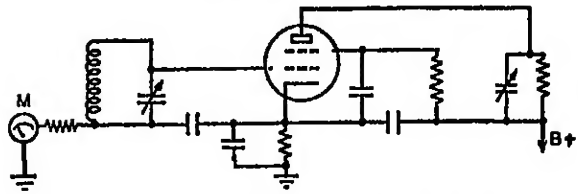
Tests.[1] The former series includes the *Purdue Test for Electricians,* the *Purdue Test for Machinists and Machine Operators,* and the *Purdue Blueprint Reading Test.* The latter series now has in the process of development and standardization tests in the following areas: oxyacetylene welding, arc welding, automobile mechanics, and engine lathe operation.

Trade Tests and Upgrading.—The problem of "who shall be upgraded" is one that is the basis of many grievances and management problems. Where the union contract places the entire emphasis on seniority or when the usual phrase "employees shall be advanced to better paying operations on the basis of seniority provided they are able to do the work," management is sometimes at a disadvantage. Even under the latter type of arrangement management has found it difficult when a particular employee with high seniority did not have the ability to do the work. The use of trade or competency tests in the upgrading process seems to be one part of the answer. The Collins Radio Company is an outstanding example. This company through an agreement with the union has established trade tests for most classifications, and adequate test performance entitles the employee to a trial period on the job. According to the company's testimony, "the use of tests has reduced failures during the trial period to less than 2 per cent." [2] Below are ten questions selected from the test in basic radio theory for use in upgrading employees to the classification Test Technician B.

1. In a radio-frequency amplifier stage having a plate voltage of 1,250 volts, a plate current of 150 milliamperes, a grid current of 15 milliamperes, and a grid-leak resistance of 4,000 ohms, what is the value of the operating grid bias?
2. Sketch a block diagram of a crystal-controlled transmitter, using a buffer stage and high-level modulation.
3. Why is the distributed capacity of a coil always increased by the wax or other coating used for protection against moisture?
4. How is the vacuum-tube plate current of an RF amplifier affected as the plate-resonant frequency is varied?

[1] Copyrighted by the Purdue Research Foundation and distributed by the Division of Applied Psychology, Purdue University, Lafayette, Ind.

[2] RANKIN, BERNARD Notes on employee testing. *Collins Signal,* January, 1947, 14, 23 (Collins Radio Co., Cedar Rapids, Iowa).



5. The above is a circuit of an RF amplifier. What is the fundamental difference in the action of the meter if the tube is operating Class C from what it would be if the tube was operating Class A?
6. What are the principal output-voltage ripple frequencies in a full-wave rectifier?
7. What is the relation between the direct-current power input of the plate circuit of the stage being modulated and the output audio power of the modulator for 100 per cent sinusoidal modulations?
8. Sketch a block diagram of a superheterodyne receiver showing an audio-frequency stage, radio-frequency stage, audio power-amplifier stage, speaker, mixer, second detector, and intermediate-frequency stage.
9. What is the sum of 25 cycles, 25 kilocycles, and 25 megacycles?
10. What is a "parasitic oscillation?"

As the use of trade tests for upgrading purposes becomes more and more accepted by labor and management, it is quite possible that there will be industry-wide developments in trade tests for that purpose. Until then, however, commercial tests will be available for only the most common trade areas.

Interview Aids.—Of interest in connection with this discussion of trade tests is a series of so-called "interview aids" that have been published in the Purdue Vocational Series. There are three tests in this series, *Can You Read a Micrometer?*, *Can You Read a Scale?*, and *Can You Read a Working Drawing?* Each is a one-page test and requires less than ten minutes to administer. As the name implies, they are useful in employment situations where it is necessary to evaluate the applicant's skill in these simple but nonetheless basically fundamental skills.

Trade Tests as Aptitude Tests.—When is a test a trade test, and when is it an aptitude test? This often-asked question probably cannot be answered to the satisfaction of all critics. Generally speaking, a trade test is a test designed to measure learning that has accrued from specific occupational training or experience. On the other hand, although everything that any

test can measure must be learned or acquired in some fashion, in the strictest sense an aptitude test is one that measures skills or knowledges which have been more or less inadvertently acquired or are a function of maturity. More generally, however, the term *aptitude* is used to denote any test that is useful in predicting job or training success. For example, the *Purdue Industrial Training Classification Test* discussed in Chap. IV is technically an achievement test. It measures one's ability to apply fundamental arithmetic in practical situations. However, as indicated in the discussion regarding the selection of electrician trainees, it predicts success in the training program better than any other single test. The *Purdue Mechanical Adaptability Test*[1] discussed later is basically an inventory of an individual's store of practical mechanical and electrical information. Since it predicts success fairly well in a number of industrial jobs, it can be considered an aptitude test. The purpose of this discussion is to point out that whether a given test is or is not an aptitude test is in reality an academic question. What the personnel or employment man wants to know is "Does this test adequately evaluate an applicant's prior experience in the field?" or "Does this test help to predict success or failure of an individual who has not done this kind of work before?" In the remaining sections of this chapter no effort will be made to tag tests as *this* or *that* kind, and the position is here taken that the less one worries about whether or not Test A is an aptitude test or not the better off he is. The pragmatic question is "Does it work?"

ASSEMBLERS, PACKERS, AND INSPECTORS

Watch Assemblers.—Blum[2] used O'Connor's *Tweezer Dexterity Test* and *Finger Dexterity Test* in studying a group of 152 female assemblers in a watch factory. Using turnover as a criterion and establishing critical scores on each of the two tests, he was able to demonstrate a significant relationship. Figure 9-2

[1] LAWSHE, C. H., JR., SEMANEK, IRENE A., AND TIFFIN, JOSEPH. The Purdue mechanical adaptability test. *J. appl. Psychol.*, 1946, 30, 442-453.

[2] BLUM, MILTON L. A contribution to manual aptitude measurement in industry. *J. appl. Psychol.*, 1940, 24, 381-416.

from Blum's data shows the percentage of each tenure group who passed both tests. Whereas only 22 per cent of those who stayed on the job a week or less exceeded the critical score on both tests, 67 per cent of those who remained on the job a year or more performed above the critical score level on both tests. Blum's study further demonstrated the relationship between performance on these two tests and job success as measured by the ratings of supervisors. An earlier study by Blum and Candee [1] was less conclusive.

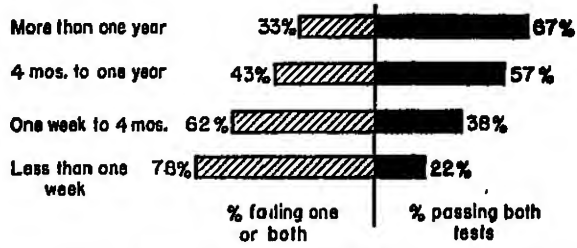


FIG. 9-2.—Proportion of employees in various tenure groups who met or failed to meet the critical score on two dexterity tests. (*From Blum.*)

Electrical-fixture and Radio Assemblers.—In research designed to select tests for identifying potentially good assemblers of electrical fixtures and radios, Tiffin and Greenly [2] studied three different job classifications. While they found different tests to be useful in the various classifications, each of the following was useful in one or more of the classifications: finger-dexterity test, hand-precision test, intelligence test, and vision tests. In two of the classifications they found height and weight to be indicative of success. In one classification, that of radio assembler, a combination of finger dexterity, hand precision, near visual acuity, and color vision yielded a multiple correlation of 0.60 with efficiency ratings. In a recent study involving radio-assembly operators Goodman [3] has reported significant relationships between selected subtests in the *MacQuarrie Test for Mechanical Ability*

[1] CANDEE, BEATRICE, AND BLUM, MILTON L. Report of a study done in a watch factory. *J. appl. Psychol*, 1937, 21, 572–582.

[2] TIFFIN, JOSEPH, AND GREENLY, R. J. Employee selection tests for electrical fixture assemblers and radio assemblers. *J. appl. Psychol.*, 1939, 23, 240–263.

[3] GOODMAN, CHARLES H. The MacQuarrie test for mechanical ability: I. *Selecting radio assembly operators. J. appl. Psychol.*, 1946, 30, 586–595.

and estimates of training success. He reports selection effectiveness as being about 12 per cent better than chance.

Radio Tube Mounters.—Forlano and Kirkpatrick report an investigation [1] involving twenty radio-tube mounters who were administered the *Otis Self-administering Test of Mental Ability* and two temperament tests, Washburne's *Social Adjustment Inventory* and the *Bell Adjustment Inventory.* By a weighting scheme in which equal value was given to the mental ability test and the two temperament tests combined, a marked relationship between test scores and supervisory ratings was demonstrated. They conclude that although a person of low intelligence is likely to fail or do poorly on the job, high mental ability does not ensure success. The two temperament scales when added to the mental ability test materially improve the "batting odds" in predicting success as measured by supervisory ratings.

Glove Assemblers.—Blum [2] in another study of assemblers in a glove factory devised the *Blum Sewing Machine Test* which is scored in number of seconds with error allowances. Dealing with experienced operators engaged in "forchetting, thumb inserting, closing and pointing," he found that various critical scores on the test eliminated larger percentages of successful operators than unsuccessful ones.

Perhaps logically belonging in the next section but so closely related that it is included here is Shartle's study [3] involving seventy-seven power-sewing-machine operators. Using a number-comparison test, a names-comparison test, and four parts from the *MacQuarrie Test for Mechanical Ability,* he demonstrated a significant relationship between combined scores and job success. Figure 9-3 shows that of those rated as most proficient, average, and least proficient, 86, 76, and 38 per cent, respectively, were in the upper two-thirds on test scores.

[1] Forlano, George, and Kirkpatrick, Forrest H. Intelligence and adjustment measurements in the selection of radio tube mounters. *J. appl. Psychol.,* 1946, 30, 257–261.

[2] Blum, Milton L. Selection of sewing machine operators. *J. appl. Psychol,* 1943, 27, 35–40.

[3] Shartle, C. L. *Psychological aids in the selection of workers.* Personnel Series No. 50. New York: American Management Association, 1941.

Food Canners.—Benge [1] made a study of 173 food packers in a cannery in which he used a manipulative test involving the placing of disks or pegs. He divided the group into seventy-nine who were most satisfactory and ninety-four who were least satisfactory on the basis of merit-rating scores and selected a critical

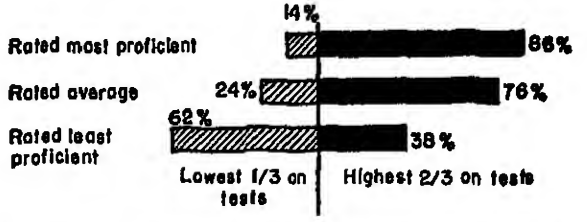


FIG 9-3.—Proportion of each of three ability groups of power-sewing-machine operators who were among the upper two-thirds in test performance. (*From Shartle.*)

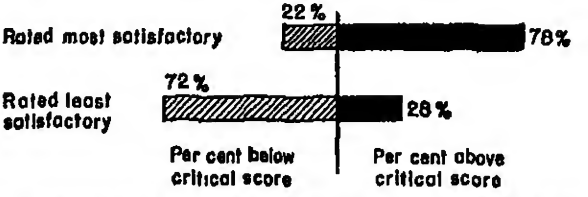


FIG. 9-4.—Proportions of the most satisfactory and least satisfactory groups of food packers who were above the critical score on a hand-dexterity test. (*From Benge.*)

score on the test. Figure 9-4 plotted from his data indicates that whereas 78 per cent of those rated as most satisfactory exceeded the critical score, only 28 per cent of those rated least satisfactory were above the critical score.

Pharmaceutical Packers.—Ghiselli [2] studied twenty-six inspector-packers in a pharmaceutical house whose job consisted of the following:

1. Filling capsules, vials, and bottles with serums, antitoxins, and similar biologicals
2. Stoppering the containers

[1] BENGE, EUGENE J. Use tests in selecting personnel. *Food Packer*, December, 1944, 36–37.

[2] GHISELLI, EDWIN E. Tests for the selection of inspector-packers. *J. appl. Psychol.*, 1942, 26, 468–476.

3. Examining them for the presence of extraneous foreign material
4. Labeling them
5. Cartoning and packaging them

Using the combined ratings of the supervisor and the floorlady as a criterion, he secured coefficients of correlation with scores of tests as follows: *Minnesota Paper Formboard,* 0.57; *Minnesota Rate of Manipulation Test* (turning), −0.40; a pegboard, −0.50.

Industrial Inspection.—The inclusion of the assembler-inspector job discussed above makes some observation on industrial inspection pertinent. The word *inspector,* as every informed personnel man knows, is used to apply to many classifications. Sometimes an inspector traces down defects in the wiring of a radio; sometimes he simply uses go and no-go gauges; and sometimes he passes or rejects material on the basis of appearance alone. From the standpoint of personnel testing, the third classification of visual inspection is one of the most important areas and is treated in Chap. VIII. So diversified, however, is inspection that a section has not been specifically assigned to it. Inspection is so often combined with other operations or is so closely related to other jobs that to emphasize the inspection phase is inappropriate. Tiffin and Rogers' [1] study involving assorting-room operators in a tin-plate mill is illustrative. They established scores on three vision tests and on the *Purdue Hand Precision Test* and, in addition, established a minimum weight of 118 pounds and a minimum height of five feet two inches. It seems obvious that height and weight have nothing to do with inspection as such. But successful girls on this job had to handle large volumes of metal in the course of the day, and physical stamina is apparently related to height and weight. At any rate, height and weight were found to be important predictors. Many other jobs in the inspection category make important demands on the individual that are not implied in the word *inspection.* In

[1] TIFFIN, JOSEPH, AND ROGERS, H. B. The selection and training of inspectors. *Personnel,* 1941, 18, 3-20.

another study of inspectors Shuman [1] found that selection of a particular type of inspector in the aircraft-engine and propeller industries could be significantly improved through the use of Bennett's *Test of Mechanical Comprehension*, the *Otis Self-administering Test of Mental Ability*, and the *Revised Minnesota Paper Form Board*.

OPERATORS AND MACHINE ATTENDERS

Aircraft-riveters and Sheetmetal Trainees.—The occupational analysis section of the United States Employment Service, in the process of developing approximately 170 test batteries, set up a three-test battery for aircraft-riveter trainees.[2] The battery consisted of a pegboard, a finger-dexterity test, and a figure-copying test. In an experimental sample of fifty-one trainees in a national defense training program the battery yielded a correlation of 0.60 with the criterion. Further study showed that the third scoring highest on the battery as compared with the third scoring lowest was 26 per cent higher in number of rivets driven per hour. Hardtke,[3] working with four groups of aircraft sheet-metal trainees, set up test batteries involving "measures of dexterity and spatial perception" and obtained multiple coefficients of correlation ranging from 0.36 to 0.65.

Coil Winders.—Hayes [4] used two pegboards, O'Connor's *Finger Dexterity Test* and the *Western Electric Pegboard*, in combination with an evaluation of prior experiences as a predictive battery for coil winders. Figure 9-5 shows the relationship between composite scores and estimates of learning speed. Cook [5] in studying the same job classification set up a manipulative test that demonstrated a significant relationship with success on the

[1] SHUMAN, JOHN T. The value of aptitude tests for factory workers in the aircraft engine and propeller industries. *J. appl. Psychol.*, 1946, 30, 156–160.

[2] Measuring occupational aptitudes. *Occupations*, 1944, 22, 387–446.

[3] HARDTKE, E. F. *Development of an aptitude test battery for aircraft sheet metal trainees.* Ph.D. thesis on file in library, University of Wisconsin, Madison, Wis., 1943.

[4] HAYES, ELEANOR G. Selecting women for shop work. *Person. J.*, 1932, 11, 69–85.

[5] COOK, D. W. *Psychological aids in the selection of workers.* Personnel Series No. 50. New York: American Management Association, 1941.

job. Figure 9-6 shows that Cook was able to establish a critical score above which 92 per cent of the better employees and only 28 per cent of the poorer employees fell.

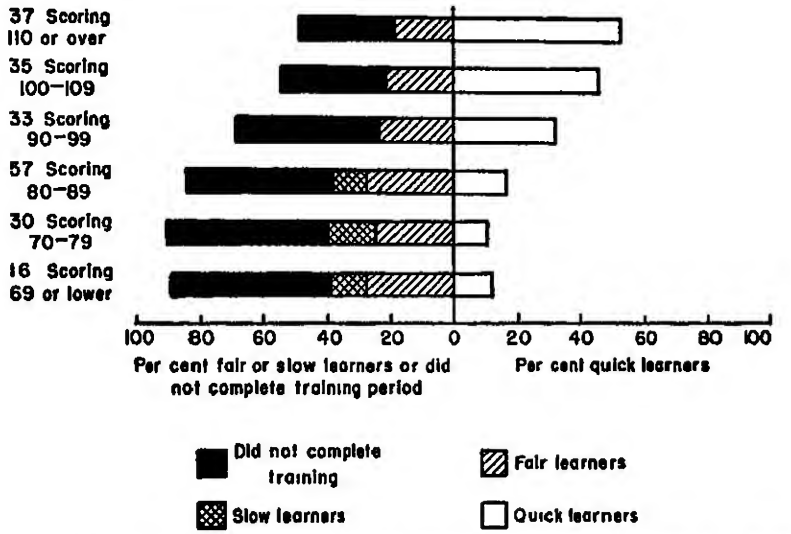


FIG. 9-5.—Relation between composite score on two dexterity tests and speed of learning for a group of 208 coil winders. (*From Tiffin*[1])

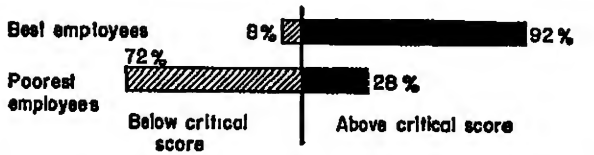


FIG. 9-6.—Proportion of best and poorest coil winders who attained and failed to attain a critical score on a manipulative test. (*From Cook.*)

Solderers.—Cook[2] reports an investigation involving solderers for whom he established a battery consisting of the *Otis Self-administering Test of Mental Ability*, an apparatus monotony test, and a finger-dexterity test. The employees were classified as above average or below average in job performance, and Fig. 9-7 shows that when he established the average score of the group

[1] TIFFIN, JOSEPH. *Industrial Psychology.* New York: Prentice-Hall, 1947, p. 136.

[2] COOK, D. W. *Psychological aids in the selection of workers.* Personnel Series No. 50. New York: American Management Association, 1941.

as his critical score, 89 per cent of the superior group on the job exceeded it whereas none of the poorer group scored above average.

Relay Adjusters.—In the same report [1] Cook presents the results of a study involving electrical-relay adjusters. He divided

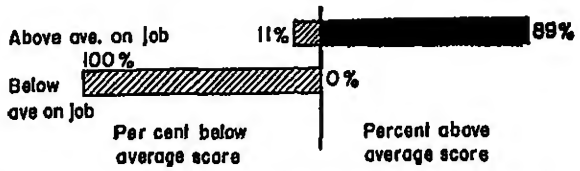


FIG 9-7.—Proportion of above- and below-average solderers who attained or failed to attain a composite critical score for a test battery. (*From Cook.*)

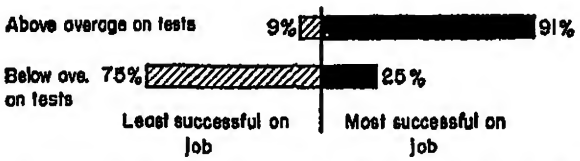


FIG. 9-8.—Proportion of relay adjustors scoring above and below average on tests who were considered most successful and least successful on the job. (*From Cook.*)

the group into the upper and lower half on their combined scores on two tests including a monotony test. Figure 9-8 indicates the relationship between combined scores and job success. Note that among those who were above average on the combined tests, 91 per cent were considered successful on the job whereas of those below average on the tests only 25 per cent were in the successful group.

Cable Formers.—Two maze tests and a tweezer dexterity test proved useful in differentiating between the more efficient and the less efficient employees engaged in a cable-forming operation.[2] Figure 9-9 shows that of those above average on test scores 80 per cent were above average in efficiency and that of those below average on test scores 25 per cent were above average in efficiency.

Bench Workers.—Job titles are frequently poor indicators of the true nature of the job, and the same title frequently is used

[1] *Ibid.*
[2] *Ibid.*

for designating a wide variety of different jobs. The title of bench hand or bench worker is no exception. Hayes[1] reports a study in which two dexterity tests in combination with an index of experience were related to rate of learning with 308 bench

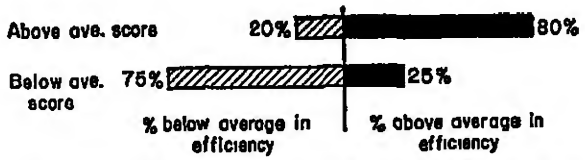


FIG. 9-9.—Proportion of above- and below-average cable formers scoring above and below average on maze and tweezer dexterity tests. (*From Cook*)

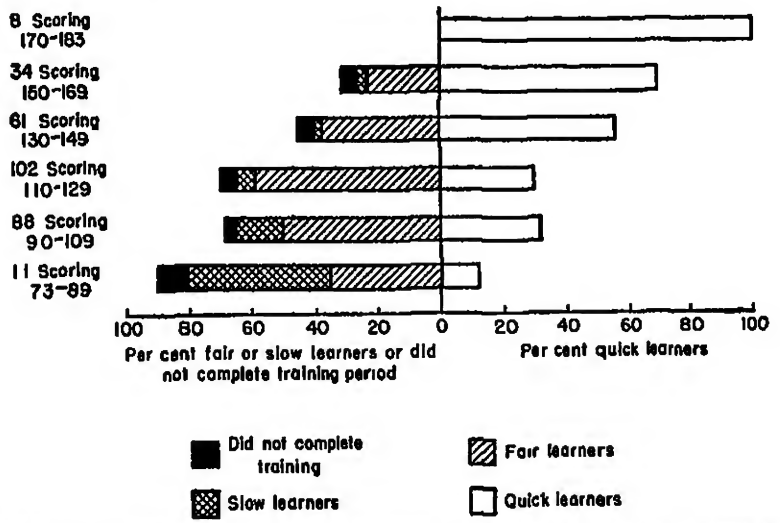


FIG. 9-10.—Relationship between composite score on two dexterity tests and speed of learning for a group of 308 bench hands. (*From Tiffin.*[2])

hands. Figure 9-10 shows that the proportion of quick learners ranged from 100 per cent among the group having highest composite scores down to 12 per cent among the group having lowest composite scores. Hayes further reports a follow-up study of sixty-two new hires. At the end of their first six months of employment the fifteen with the highest composite scores were performing considerably better on the job than the remainder.

[1] HAYES, *op. cit.*

[2] TIFFIN, *op. cit.*, p. 136.

Laundry Workers.—Although it has been the policy in this and other chapters to report only those studies in which objective validity data are presented, it seems desirable to deviate from that policy in order to cite Ten Broeck's report [1] on the selection of laundry workers. No validity data are presented, but the author reports excellent results with a battery that includes the following tests: *Minnesota Rate of Manipulation Test, Otis Employment Test, Minnesota Vocational Test for Clerical Workers,* and Washburne's *Social Adjustment Inventory.* Obviously there is a need for objective studies in the laundry industry.

Punch-press Operators.—In her study [2] previously cited, Hayes studied 254 operators of punch presses and similar machines. Here again she used the same two pegboards plus an index of prior experience. Among those in the highest bracket on the composite score 65 per cent were considered quick learners, whereas only 20 per cent of those in the lowest score bracket were so rated. The miniature punch press described by Tiffin and Greenly [3] is an example of a trade test of the manipulative variety. While ability grouping validity data are presented, the authors demonstrate that the test differentiates between groups of people with and without punch-press experience. They further found that test performance was related to speed and accuracy ratings made by foremen, the coefficients of correlation being −0.55 and 0.63. They also report a correlation of 0.60 between test scores and safety ratings.

Paper-converting-machine Operators.—Jurgensen [4] reports a study involving 212 operators of paper-converting machines. In describing the job the author indicates that the operators were engaged in removing tissues from the machine, inserting advertising matter, and placing them on conveyors. Characteristic of many machine-tending jobs, its emphasis is fundamentally on

[1] Ten Broeck, Delphine L. How aptitude tests can aid you in employee selection. *Laundry Age,* Jan. 1, 1945, 36–38.

[2] Hayes, *op. cit.*

[3] Tiffin, Joseph, and Greenly, R. J. Experiments in the operation of a punch press. *J. appl. Psychol.,* 1939, 23, 450–460.

[4] Jurgensen, Clifford E. Extension of the Minnesota Rate of Manipulation Test. *J. appl. Psychol.,* 1943, 27, 164–169.

packing. Jurgensen evaluated job adequacy by means of three supervisory ratings on each employee. Using certain extensions of the *Minnesota Rate of Manipulation Test,* he obtained a coefficient of correlation of 0.61 with these ratings.

Bar-mill Employees.—In one steel-mill study a battery of three tests was administered to each of twenty-eight applicants at the time when they were hired as bar-mill employees. Each of the tests, the *Purdue Industrial Training Classification Test,*

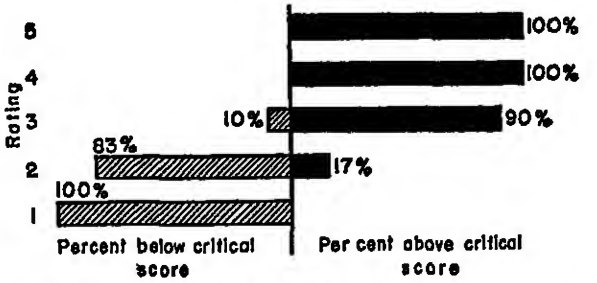


FIG. 9-11.—Proportion of each of five rated groups of bar-mill employees who scored above and below a critical score on the *Purdue Industrial Training Classification Test.*

the *Adaptability Test,* and the *Purdue Mechanical Adaptability Test,* showed significant relationships with supervisory ratings, the correlation coefficients being respectively 0.72, 0.59, and 0.57. The *Purdue Industrial Training Classification Test* was particularly effective as is demonstrated in Fig. 9-11. Note that when a critical score of 8 is imposed, 100 per cent of those rated 4 or 5 are above and all of those rated 1 and 83 per cent of those rated 2 are below.

MACHINE-TOOL LEARNERS AND APPRENTICES

Nature of Machine-tool Work.—Considerable confusion exists in the thinking of many people when they consider the requirements of jobs that utilize machine tools. Probably no other single area of work activity embraces employees with such a wide range of talent. A lathe, for example, may be operated by a routine employee engaged in a highly repetitive operation which requires an absolute minimum of planning and judgment by the

operator. Or the same lathe may be operated by a specialist, an individual who is quite expert on all phases of lathe operation, who plans and executes the lathe work including the setup activity, but who professes no skill at all on other machine tools. And, finally, a lathe may be one of many tools utilized by an expert machinist or tool- and diemaker. To classify all of these and the many others that could be named together in the testing procedure would be similar to grouping together all people who use pencils. True, the manipulative act of pushing the pencil is the

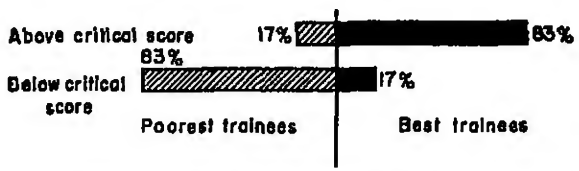


FIG. 9-12.—Proportion of machine-tool-operator trainees above and below critical test scores who were considered in the best and poorest group. (*From Ross.*)

same from job to job, but other aspects of the work tend to magnify or minimize the relative importance of pencil utilization from job to job. Likewise, the progression of jobs from the operator level to the tool- and diemaker level is such that although minimum manipulative skills are important all along the line, the *relative* importance of the manipulative phase of the total job diminishes as the job level increases. Consequently, the battery of tests that will pick good machine-screw operators will not necessarily and will probably not pick those most apt to succeed in a four-year apprenticeship program.

Operators and Trainees.—Patten [1] in an early study used a coordination test, a lock test, a box test, and a circle-centering test in a battery to forecast ability to operate a lathe as measured by proficiency in making five standard jobs on the lathe. Ross,[2] working with a group of adult trainees on various machine tools, administered O'Connor's *Finger Dexterity Test* and by estab-

[1] PATTEN, EVERETT F. An experiment in testing engine lathe aptitude. *J. appl. Psychol*, 1923, 7, 16-29.

[2] ROSS, LAWRENCE W. Results of testing machine-tool trainees. *Person. J.*, 1943, 22, 303-307.

lishing a critical score of 304 seconds was able to show a significant relationship with job performance. Figure 9-12 shows that 83 per cent of those above the cutoff were in the best category whereas only 17 per cent of those below the cutoff were considered best.

Bennett and Fear,[1] working with operators of turret lathes, precision grinders, milling machines, and Bullard automatics, found a significant relationship between composite scores on Bennett's *Test of Mechanical Comprehension* and a hand-tool

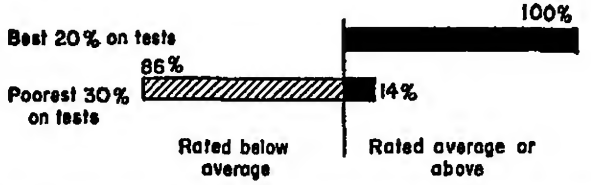


FIG. 9-13.—Proportion of machine-operator groups scoring high and low on a test battery who were rated above and below average. (*From Bennett and Fear*)

dexterity test and ratings. As is demonstrated in Fig. 9-13, of those scoring in the upper 20 per cent on the two tests 100 per cent were considered average or above on the job whereas of those scoring in the lowest 30 per cent on the tests only 14 per cent were considered average or above. The authors further state that new men who were hired after the battery was installed were rated consistently higher than was the case prior to testing and that not a single new man hired since tests were introduced as part of the selection procedure has had to be dismissed because of lack of ability to do the job.

Numerous English studies have dealt with the so-called "engineering operatives," an example of which is Andrews's study.[2] She administered to 122 miscellaneous machine-shop workers' tests designed to measure the following: intelligence control of movement, steadiness of movement, finger dexterity, accuracy of placing, bimanual coordination, and observation.

[1] BENNETT, GEORGE K., AND FEAR, RICHARD A. Mechanical comprehension and dexterity. *Person. J*, 1943, 22, 12–17

[2] ANDREWS, AMY C. A year's experience of selection tests for engineering operatives. *Occup. Psychol.*, 1944, 18, 126–130.

She reports that whereas 60 per cent of the whole group were rated as satisfactory, of the half scoring highest on the test that had been accepted 98 per cent were rated satisfactory.

One test that shows promise is the *Purdue Mechanical Assembly Test.*[1] This is an individual apparatus test involving mechanisms that are new to all testees. The author has reported validity coefficients as high as 0.55 between supervisory ratings and test scores of machinists and machinist's helpers.

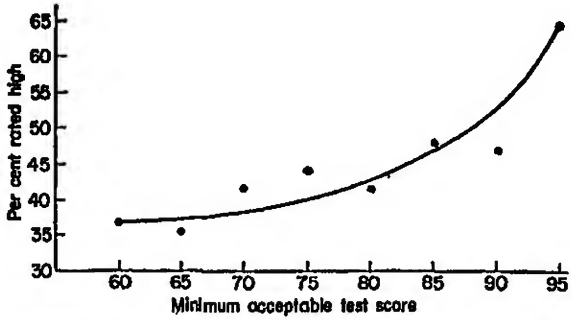


FIG. 9-14.—Proportion of a group of machine operators rated high when various critical scores are employed. (*From Lawshe, Semanek, and Tiffin.*)

Screw-manufacturing Employees.—One of the newer tests that has been useful in identifying inexperienced machine operators who are most apt to make good on the job is the *Purdue Mechanical Adaptability Test.* The test, which is really an inventory of experience in mechanical, electrical, and related areas, was used with some success in connection with a group of forty-six operators in a plant engaged in the manufacturing of screws. The operators were rated by supervisors, and Fig. 9-14 from a study by Lawshe, Semanek, and Tiffin [2] shows the results. While approximately 37 per cent of the total group were rated high, the proportion gradually increases as higher critical scores are applied. When the critical score of 85, for example, is applied, the proportion of high rated employees increases to about 48 per cent.

[1] GRANEY, M. R. *The construction and validation of a new type of mechanical assembly test.* Ph.D. thesis, Purdue University, 1942.

[2] LAWSHE, C. H., JR., SEMANEK, IRENE, AND TIFFIN, JOSEPH. The Purdue Mechanical Adaptability Test. *J. appl. Psychol.*, 1940, 28, 442-453.

Tool Setters.—When there is a high degree of specialization, the tool setter becomes a key employee. Crissey [1] reports a study in which he used a battery consisting of the *Minnesota Spatial Relations Test,* a pegboard, and a peg-turning test. Figure 9-15 shows the relationship that he obtained between composite scores and supervisory ratings. Note that among the one-third scoring highest on the tests, 69 per cent were in the third rated highest and that none were in the third rated lowest. Among

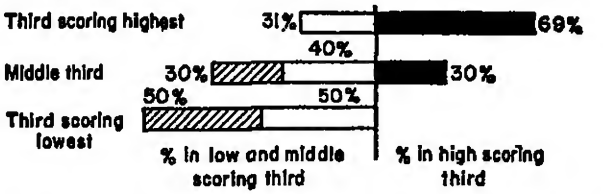


FIG. 9-15.—Proportion of three rated ability groups of tool setters who made test scores in the high third (solid bar), middle third (open bar), and lower third (shaded bar) of the group. (*From Crissey.*)

the third scoring lowest on the tests, 50 per cent were in the third rated lowest and none were in the third rated highest. Shuman [2] in his aircraft-engine and propeller-plant study showed significant improvement in the selection of tool setters by means of the *Otis Self-administering Test of Mental Ability,* Bennett's *Mechanical Comprehension Test,* and the *Revised Minnesota Paper Formboard.*

Machinist Apprentices.—Pond [3] in making a general evaluation of a testing program for machinist apprentices compared the supervisory ratings of 163 apprentices who were hired without the use of tests and another 155 who were selected after testing had been added to the selection procedure. Figure 9-16 shows her results. Without tests, 61 per cent were considered superior; but when tests were used, 83 per cent were rated in the superior category. With respect to the type of apprentice selected she

[1] CRISSEY, ORLO L. *The use of tests in improving personnel procedures.* Flint, Mich.: General Motors, 1944. 24 pp.

[2] SHUMAN, *op. cit.*

[3] POND, MILLICENT. What is new in employment testing. *Person. J.,* 1932, 11, 10-16.

says that "adoption of the minimum scores for tool-making apprentices has improved the quality of the group selected as much as was formerly accomplished in a year of trial in the course." When Shuman [1] applied his battery consisting of the *Otis Self-administering Test of Mental Ability*, Bennett's *Mechanical Comprehension Test*, and the *Revised Minnesota Paper Form-*

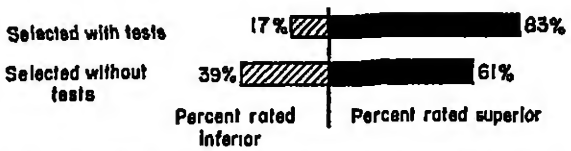


FIG. 9-16.—Proportion of machine apprentices selected with and without tests who were rated inferior and superior. (*From Pond.*)

board to a group of toolmaker learners he again obtained significant results, the Bennett showing the highest relationship with the criterion of any of the three.

SERVICE ELECTRICIANS AND REPAIRMEN

Electrical Troublemen.—Electrical troublemen, or trouble shooters, who "make quick and temporary repairs on any break that may occur in the power or lighting transmission circuits in the metropolitan area" were studied by Shartle.[2] His battery included, among others, the pursuit test and the blocks test from the *McQuarrie Test for Mechanical Ability*, an arithmetic test, an electrical-circuit test, and an electrical information test. The troublemen were rated by supervisors and placed in A, B, and C classes. By establishing a critical score Shartle was able to obtain the results presented in Fig. 9-17. Note that whereas of those rated C only 25 per cent were above the critical score, of those rated A all, or 100 per cent, were above the critical score. Also significant is the fact that when he classified these same troublemen into two groups, one composed of those who had been involved in one or more accidents during a stipulated period and the other composed of those who had not, 67 per cent of the for-

[1] SHUMAN, *op. cit.*

[2] SHARTLE, CARROLL L. A selection test for electrical troublemen. *Person. J.*, 1932, 11, 177–183.

mer were above the critical score as compared with 80 per cent of the latter. Although the difference is not large, it is sufficiently important to be considered.

Dial Switchmen.—The job title, dial switchman, is used to designate maintenance men in a dial switchboard telephone office. Rossett and Arakelian [1] studied a group of these switchmen and

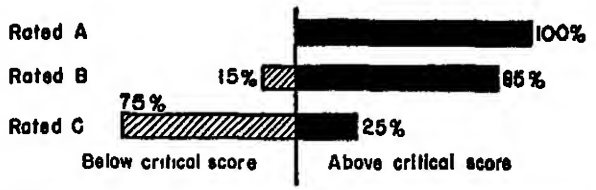


FIG. 9-17.—Proportion of three rated ability groups of electrical troublemen who attained or failed to attain a given critical score. (*From Shartle.*)

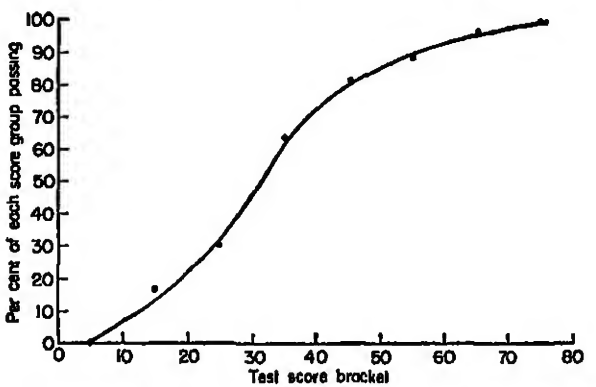


FIG. 9-18.—Proportion of dial switchmen trainees in each test score bracket who passed training course. (*From Rossett and Arakelian.*)

developed a fifty-minute test battery of ten tests including the following: electrical information, knowledge of diagrams and apparatus, adjustment of apparatus, and some maze tests. All switchmen attended a training school, and the authors obtained a correlation of 0.68 between test scores and rated success in the school. Perhaps even more significant is an analysis of the actual failures in the school. During the time of the investigation a

[1] ROSSETT, NATHANIEL E, AND ARAKELIAN, PETER. A test battery for the selection of dial switchmen. *J. appl. Psychol.*, 1939, 23, 358-366.

total of 222 men were involved of whom 43 were transferred to other departments because of failure. Figure 9-18 presents an analysis of these failures by test scores. Of those scoring from 10 to 20, 17 per cent passed; or, in other words, 83 per cent were transferred to other work. Notice that the curve moves steadily upward and that among those scoring between 70 and 80, 100 per cent passed.

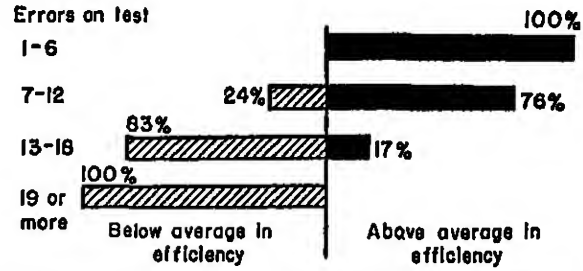


FIG. 9-19.—Proportion of electrical testers and inspectors scoring at various levels on a perception test who were above average and below average on job efficiency. (*From Cook.*)

Electrical Testers and Inspectors.—Cook[1] in the report referred to used a perception test in a study involving a group of electrical testers and inspectors whose work consisted primarily of tracing wiring diagrams and locating trouble. Figure 9-19 indicates that when the employees were divided into the half above average in efficiency and the half below average in efficiency, a significant relationship with errors on the test was revealed. The figure shows that of those with six or fewer errors 100 per cent were in the upper half and of those having nineteen or more errors 100 per cent were in the lower half.

Ice-company Mechanics.—The *Purdue Mechanical Adaptability Test* previously mentioned was administered to fourteen general mechanics, or "handy men," in an ice company.[2] These same men were also rated on a five-point scale by the owner-manager. Figure 9-20 shows the results in a scattergram form. Not only did the highest rated man receive the highest score and

[1] COOK, *op. cit.*

[2] LAWSHE, SEMANEK, AND TIFFIN, *op. cit.*

the lowest rated man receive the lowest score, but most others tended to fall in line. Note, for example, that a critical score of 95 would pass all of those rated 4 or 5 and fail all of those rated 1 or 2.

Aircraft Mechanic Learners.—Jacobsen [1] reports a study involving several different classifications of aircraft mechanic learners which are grouped together here for convenience. All

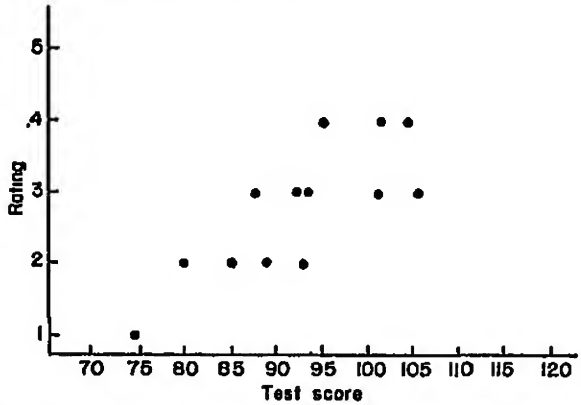


FIG. 9-20.—Relationship between efficiency rating and *Purdue Mechanical Adaptability Test* score for fourteen ice-company mechanics. (*From Lawshe, Semanek, and Tiffin.*)

learners were in training classes, and correlations between test scores and class proficiency are given. With aircraft instrument repairmen he used Pressey's *Senior Classification Test,* Pressey's *Senior Verifying Test,* and Bennett's *Test of Mechanical Comprehension* and obtained a multiple correlation of 0.61. Using the same battery with a group of aircraft repair mechanics, he obtained a correlation of 0.42. For the classification of aircraft electrician he used Bennett's *Test of Mechanical Comprehension* and a test designed to measure three-dimensional visualization.

Cotton-mill-machine Fixers.—Using an adaptation of the *Stenquist Mechanical Assembly Test,* Harrell [2] had forty-five

[1] JACOBSEN, ELDON E. An evaluation of certain tests in predicting mechanic learner achievement. *Educ. Psychol. Meas.,* 1943, 3, 259–267.

[2] HARRELL, WILLARD. The validity of certain mechanical ability tests for selecting cotton mill machine fixers. *J. soc. Psychol.,* 1937, 8, 279–282.

loom fixers in a cotton mill rated by their overseers and obtained a correlation of 0.45 between scores and ratings. In another study he had the carding department overseer rate "according to mechanical ability" forty men who had been fixers and obtained a correlation of 0.84 between scores and ratings. A similar investigation with ten spring-frame fixers gave a correlation of 0.78.

OTHER TRADE GROUPS

Printing Pressmen Apprentices.—The *Minnesota Paper Formboard* was used by Hall [1] for identifying superior printing

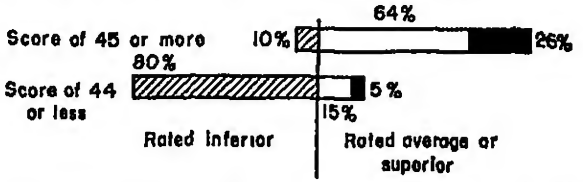


FIG. 9-21.—Proportion of printing pressmen apprentices scoring above and below a critical score who were rated inferior (shaded bars), average (open bars), and superior (black bars). (*From Hall.*)

pressmen apprentices. Eighty-six men were rated by three different instructors, and the resulting composites were divided into inferior, average, and superior. A critical score of 45 on the test was established, and Fig. 9-21 presents the results. Of those above the critical score only 10 per cent were considered inferior, and of those below the critical score 80 per cent were considered inferior. Of the first group 26 per cent were superior, and of the latter group only 5 per cent were so considered.

Miscellaneous Shop Workers.—Earle [2] reports the use of the *Stenquist Mechanical Assembly Test* with trainees in a number of shop areas. Using instructors' estimates of proficiency as a criterion, he obtained correlations as follows: fitters, 0.49; carpenters, 0.82; blacksmiths, 0.37; and electricians, 0.07. None of the groups included more than twenty-one men.

[1] HALL, MILTON O. An aid to the selection of pressmen apprentices. *Person. J.*, 1930, 9, 77–81.

[2] EARLE, F. M. *Tests of mechanical ability.* Studies in Vocational Guidance, Report No. 3. London: National Institute of Industrial Psychology, 1929. 42 pp.

Lawshe [1] demonstrated the usefulness of the *Purdue Industrial Training Classification Test* in identifying successful miscellaneous shop trainees in three different situations. The 25 per cent rated highest in every instance made significantly higher average scores on the test than did the 25 per cent rated lowest.

Mixed Apprentice Groups.—Two studies [2] have been conducted with the *Purdue Mechanical Adaptability Test* which deal with mixed apprentice groups. Although both of these groups include machinist's apprentices and consequently might have

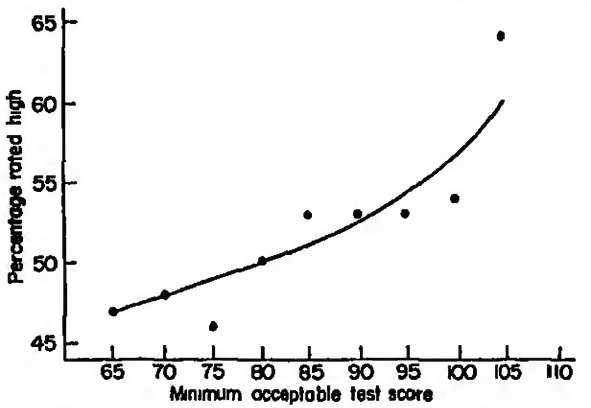


FIG. 9-22.—Proportion of apprentices who were rated high when various critical scores on the *Purdue Mechanical Adaptability Test* were considered. (*From Lawshe, Semanek, and Tiffin.*)

been discussed in an earlier section, consideration has been postponed until this point because combinations of occupational classifications are involved. One group of twelve in a steel mill included the following: four machine shop, two masonry, and one each of electrical, maintenance, carpenter shop, blacksmith, pipe shop, and electrical construction. A low but positive correlation of 0.39 between ratings and scores was found.

The other apprentice study was conducted in an electrical manufacturing plant with a group that included machinists, toolmakers, diemakers, foundrymen, and miscellaneous electrical

[1] LAWSHE, C. H., JR. The Purdue Industrial Training Classification Test. *J. appl Psychol.*, 1942, 26, 770–776.

[2] LAWSHE, SEMANEK, AND TIFFIN, *op. cit.*

workers. These men were rated by their supervisors, and Fig. 9-22 shows the results after the ratings were corrected for differences among judges. The apprentices were divided into the lowest rated 47 per cent and the highest rated 53 per cent, the latter group being called "high." As successively higher critical scores are applied, the percentage of high rated apprentices systematically increases.

SUMMARY

The term *mechanical aptitude* has little meaning, since it is employed in such a wide variety of situations. It is quite likely that there are three components of mechanical ability: the capacity to understand mechanical relationships and processes, manipulative ability, and a cluster of such motor abilities as strength, speed of movement, and endurance. Mechanical ability tests may be classified as paper-and-pencil vs. apparatus tests, or they may be classified by function into one of the following: information, spatial relations, mechanical comprehension, or interest.

CHAPTER X

TESTS FOR CLERICAL AND OTHER OFFICE EMPLOYEES

Clerical tests are among the most widely used of all tests in business and industry. Some of the reasons are obvious: (1) Clerical tests are usually of the trade test variety, frequently embodying a job sample, and have high "face validity"; (2) results are frequently more obvious in the office than in the plant, particularly when no validity studies are undertaken; and (3) office managers are frequently more easily "sold" on testing programs than are plant superintendents and industrial supervisors. In spite of the wide use of such tests, however, the literature is more void of actual validity studies than is the case with studies involving plant-operating jobs. No doubt this fact is associated with the first reason given above; they so closely resemble the real thing that it hardly seems necessary to many users to conduct validity studies.

Nature of Office Work.—Like job classifications in the plant, classifications in the office vary considerably from company to company. Generally, however, there are four types of workers: clerks, engaged in filing, checking, tabulating, and miscellaneous office tasks; typists, engaged primarily in typing but usually performing some function performed by clerks; stenographers, usually employing shorthand and typing in their work; and machine operators who may use any kind of office equipment from the duplicator or dictation transcriber to machine bookkeeping and posting equipment. Only in very large offices is there a clear demarcation among these various types of jobs, and one frequently finds such job titles as clerk-secretary and receptionist-typist. Also the extent to which the employee engages in true secretarial activity merits attention. Some employees are given

considerable responsibility and exercise ingenuity and independent judgment; others perform tasks just as routine as does any line employee in the plant. These facts should demonstrate to the employment man the necessity for validity studies on office jobs.

Criteria in Office Jobs.—One comment is frequently offered when the subject of validity in connection with office jobs is discussed: "No two girls do the same thing." This is often true. The Committee on Tests of the Life Office Management Association [1] has offered the following list of possible criteria:

1. *Salary.* (This, of course, has certain limiting factors including economic conditions, unequal opportunities from department to department, and the fact that salary may be related to length of service and not to adequacy.)
2. *Job level* or grade in the salary evaluation or classification plan.
3. *Quality and quantity* of output (in such areas as card punching, ediphone or dictaphone transcription, and the typing of standard forms or reports).
4. *Training time* (applicable only when an employee may progress at her own rate).
5. *Promotability* measured in terms of job level attained within a specified period of time.
6. *Ratings* of supervisors or others.

OFFICE CLERKS

Routine Clerical Work.—Roberts and Ostermick [2] conducted a study involving twenty-two clerks in which they used the Wonderlic *Personnel Test* in combination with the *Minnesota Vocational Test for Clerical Workers* and by using a critical score of 15 on the former and 205 on the latter were able to show sig-

[1] *The application of psychological tests to the selection, placement, and transfer of clerical employees.* Life Office Management Association, Committee on Tests. Report No. 6. New York, 1942. 28 pp. (Processed.)

[2] ROBERTS, WILLIAM H., AND OSTERMICK, RALPH E. *Test scores and ratings of Ditto machine operators.* Milwaukee, Wis.: Allis Chalmers Mfg. Co., 1945. (Mimeo.)

nificant relationships with supervisory ratings. Figure 10-1 shows their results. Note that 100 per cent of those rated above average passed both critical scores whereas only 20 per cent of those rated below average exceeded both critical scores. They report correlation coefficients with supervisory ratings of 0.54 with the Wonderlic and 0.53 with the Minnesota.

Mental Ability and Clerical Jobs.—As in the above case, mental ability tests quite frequently "come through" in various kinds

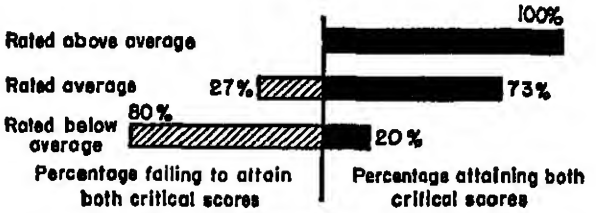


FIG. 10-1.—Proportion of clerks in three rated ability groups who attained and failed to attain critical score in a test battery. (*From Roberts and Ostermick.*)

of office and clerical jobs. Since mental ability is discussed at length in Chap. V, one study is sufficient at this point (see particularly Fig. 5-9). Pond and Bills[1] report a study involving 286 men engaged in clerical and related occupations. The *Scoville Mental Ability Test* was used, and their results are presented in Fig. 10-2. As indicated in the figure, of those scoring 170 or above 90 per cent were considered satisfactory and 10 per cent unsatisfactory. Note that the percentage considered satisfactory gradually decreases with lower score brackets.

Test Batteries.—Further insight into the nature of clerical ability is provided by two studies, each involving a number of different tests. Armstrong[2] administered a battery of six tests to a group of office and clerical employees. Included were tests designed to measure memory and attention, arithmetic, correct copy work, checking errors, filing, and common sense and reason-

[1] POND, MILLICENT, AND BILLS, MARION A. Intelligence and clerical jobs; two studies of relation of test score to job held. *Person. J.*, 1933, 12, 41–56.

[2] ARMSTRONG, T. O. New methods in promotion and hiring. *Person. J.*, 1936, 15, 280–283.

ing. He obtained a correlation of 0.80 with the ratings of departmental managers.

As a part of the Minnesota studies, Dvorak [1] presents group profiles of a group of poor female office clerks and a group of good clerks. Every one of the following tests discriminated in favor of the good employees: Pressey's *Senior Classification Test* and *Senior Verification Test, Minnesota Vocational Test for Clerical Workers,* O'Connor's *Finger Dexterity Test* and *Tweezer Dex-*

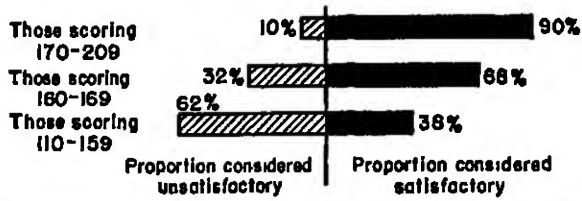


FIG. 10-2.—Proportion of three groups of clerks classified according to test scores who were considered satisfactory and unsatisfactory. (*From Pond and Bills.*)

terity Test, Minnesota Rate of Manipulation Test, Minnesota Mechanical Assembly Test (Box A), and the *Minnesota Spatial Relations Test.*

TYPISTS

Office Typists.—Cook [2] reported one of the few published validity studies involving typists. Using the follow-up method, he administered a speed and accuracy typing test to 190 typists at the time of employment. After the first six months on the job, a comparison was made between test scores and job performance. Using a "67 per cent bogey" to indicate that level of job performance below which a girl could not fall and still be called satisfactory, he obtained significant results which are shown in Fig. 10-3. Of those girls who type forty words per

[1] DVORAK, BEATRICE. Differential occupational ability pattern. *Univ. Minn. Bull. Employt. Stab. Res. Inst.*, Vol. III, No. 8, University of Minnesota Press, 1934.

[2] COOK, D. W. Some practical results from tests. *Proc. Personnel Selection Clinic.* Kansas City, Mo.: Greater Kansas City Committee for Economic Development, 1945.

minute or better on the test at the time of employment 100 per cent made the 67 per cent bogey. Of those who typed fewer than forty words per minute on the test only 45 per cent made the bogey and 55 per cent did not.

Typing Tests.—There are a number of typing tests on the market. Among the most widely used are the *Thurstone Examination in Typing* and another by Blackstone. One of the newer tests is *Kimberly-Clark Typing Ability Analysis,* in one report [1]

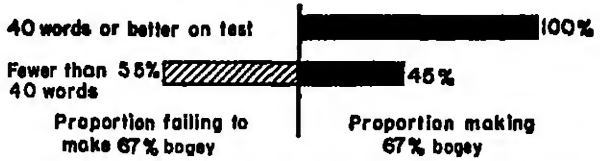


FIG. 10-3.—Proportion of typists classified according to test score who attained or failed to attain a certain job-proficiency level. (*From Cook.*)

on which are given validity correlations in the 90's between scores and estimates of adequacy of typists in some paper mills.

STENOGRAPHERS

Mental Ability.—Roberts and Ostermick [2] report a study involving sixteen stenographers in which they used only Wonderlic's *Personnel Test.* A critical score of 20 on the test differentiated between high and low rated employees as is shown in Fig. 10-4. Note that of those rated "above average" 100 per cent passed the critical score and that of those rated "below average" 100 per cent failed the critical score.

Transcription Tests.—Much testing of shorthand skill is done in a hit-or-miss fashion. Frequently standardized tests are not used; and even when the transcription section of the *Blackstone Stenographic Proficiency Test* is used, considerable skill is needed by the examiner to ensure proper timing. The Seashore and Bennett *Stenographic Proficiency Tests* which are available on phonograph records seem to be the answer. Five letters of vary-

[1] JURGENSEN, CLIFFORD E. A test for selecting and training industrial typists. *Educ. Psychol. Meas.,* 1942, 2, 409–425.

[2] ROBERTS AND OSTERMICK, *op. cit.*

ing lengths and dictated at five different speeds are included in each of two forms.

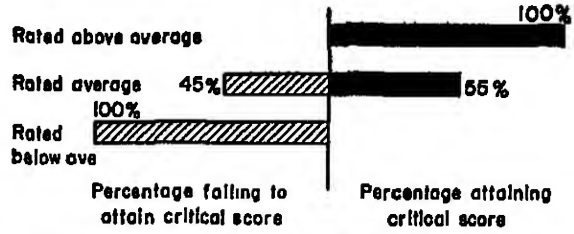


FIG. 10-4.—Proportion of three rated ability groups of stenographers who attained and failed to attain a critical score on the *Personnel Test*. (*From Roberts and Ostermick.*)

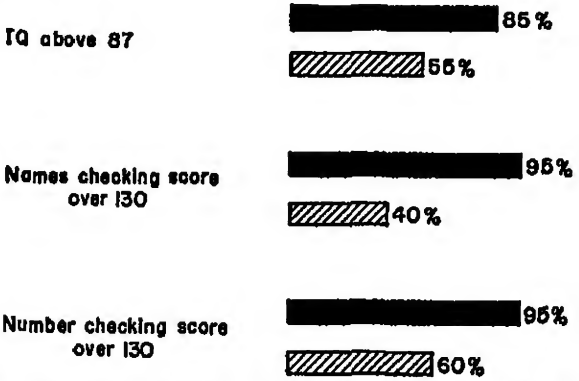


FIG. 10-5.—Proportion of better group (solid bars) and poorer group (shaded bars) of machine bookkeepers who attained critical scores on three different tests. (*From Hay.*)

OFFICE MACHINE OPERATORS

Machine Bookkeepers.—Hay[1] reported a study involving forty machine bookkeepers. He used a mental ability test and both the names and number sections of the *Minnesota Vocational Test for Clerical Workers*. When the bookkeepers were divided into the best twenty and the poorest, the results presented in Fig. 10-5 were obtained. When a critical score of 88 (I.Q.) on the mental ability test was used, 85 per cent of the better group and 55 per cent of the poorer group exceeded it.

[1] HAY, EDWARD N. Tests in industry. *Person. J.*, 1941, 20, 3-15.

More marked differences were obtained with the names and numbers checking tests as is shown in the figure.

Ditto-machine Operators.—A study by Roberts and Ostermick,[1] though involving only twelve ditto-machine operators, should be cited. Using the same critical scores indicated above in their study of clerks, they found that Wonderlic's *Personnel Test* and the *Minnesota Vocational Test for Clerical Workers* showed a significant relationship with rated success. Figure 10-6

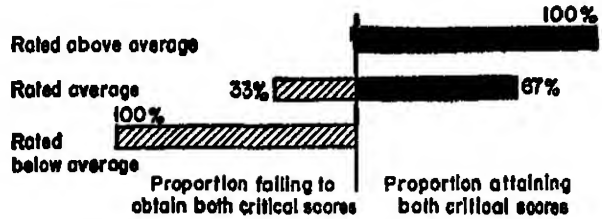


FIG. 10-6.—Proportion of three rated ability groups of ditto-machine operators who attained or failed to attain critical scores on two tests. (*From Roberts and Ostermick.*)

shows that whereas 100 per cent of those rated above average attained both critical scores, 100 per cent of those rated below average failed to attain both critical scores.

Miscellaneous Machine Operators.—Ghiselli[2] through the United States Employment Service used the *Minnesota Vocational Test for Clerical Workers* in connection with a number of different clerical job classifications, five of which involved machine operation. The number of people ranged from 26 adding-machine operators to 121 card-punch-machine operators. Figure 10-7 shows his results. Of those who were above average on the criterion the respective proportions that were above and below average on the tests are presented. Although the degree of discrimination is slight in all jobs, the consistency of the pattern indicates the almost universal presence in machine clerical jobs of the trait or skill measured by the test.

[1] ROBERTS AND OSTERMICK, *op. cit.*

[2] GHISELLI, EDWIN E. A comparison of the Minnesota Vocational Test for Clerical Workers with the general clerical battery of the United States Employment Service. *J. appl. Psychol.*, 1942, 26, 75-80.

Gottsdanker [1] worked with forty-four female learners engaged in the operation of crank-driven calculators. He used a number dot-location test which he referred to as a paper keyboard test and obtained a correlation of 0.49 with his criterion. An arith-

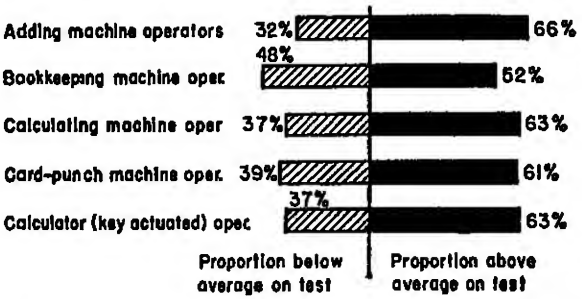


FIG. 10-7.—Proportion of those judged above average in five job classifications who were above average on the *Minnesota Vocational Test for Clerical Workers.* (*From Ghiselli.*)

metic computation test and a number-comparison test yielded correlations of 0.36 and 0.29 respectively.

JOB AND FACTOR ANALYSIS

Primary Mental Abilities.—The concept of primary mental abilities was discussed in Chap. V. An application of this concept to clerical work has been made by the Committee on Tests of the Life Office Management Association.[2] Three separate tests were designed, paralleling three of Thurstone's primary mental abilities, *N* (numerical facility), *V* (verbal facility), and *M* (memory). The first test (*N*) was administered to 113 clerical employees who were engaged in checking and posting operations. The results of the study are presented in Fig. 10-8. The range of scores was divided into three brackets, and the proportion of those scoring in each bracket who were considered above average, average, and below average in job performance is shown. Where-

[1] GOTTSDANKER, ROBERT M. Measures of potentiality for machine calculation. *J. appl. Psychol.*, 1943, 27, 233-248.

[2] *The application of psychological tests to the selection, placement, and transfer of clerical employees.* Life Office Management Association, Committee on Tests, Report No. 6. New York, 1942. 28 pp. (Processed.)

as 36 per cent of the high scorers were above average, none of the low scorers were above average.

The verbal facility test was administered to thirty-three clerical employees engaged in discussion and correspondence, and

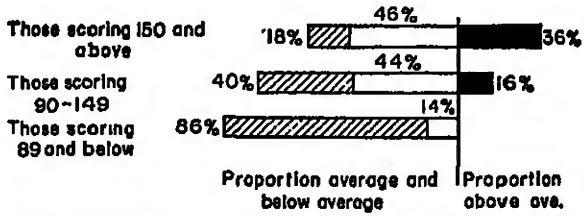


FIG. 10-8.—Proportion of checking and posting clerks classified according to test scores who were considered below average (shaded bars), average (open bars), and above average (solid bars) on the job. (*From the Life Office Management Association.*)

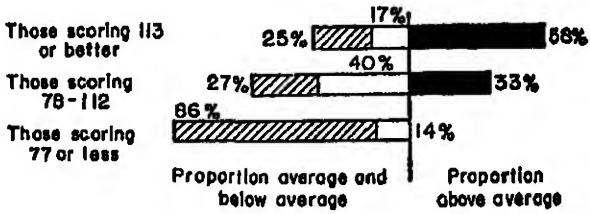


FIG. 10-9.—Proportion of clerical employees engaged in verbal activities classified according to test score who were considered below average (shaded bars), average (open bars), and above average (solid bars) on the job. (*From the Life Office Management Association.*)

similar results are presented in Fig. 10-9. Note that the proportion of above-average employees increases as the score increases and that the number of below-average employees increases as the score decreases. Figure 10-10 tells a similar story regarding Test *M* which was administered to employees engaged in typing and machine operation. Although the values vary somewhat, the pattern is identical.

Job Classification.—Another test of the Life Office Management Association identified only as Test *IA* (apparently a general mental ability test) has been used in a unique fashion that demonstrates an interesting relationship between human abilities as defined and measured by tests, and a clerical job classification

system. All clerical positions were classified into four grades as indicated below.

Grade I—Simple clerical work
Grade II—Complicated clerical work
Grade II2—More complicated clerical work
Grade III—Decision-making jobs

An analysis of the test scores of 200 clerical employees who had been on the job from five to ten years was made with regard to

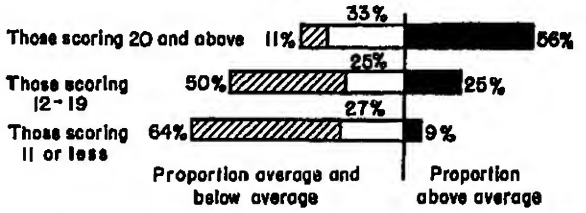


FIG. 10-10.—Proportion of typists and machine operators classified according to test score who were considered above average (solid bars), average (open bars), and below average (shaded bars) on the job. (*From the Life Office Management Association.*)

the job level or grade attained in that period of time. The results are presented in Fig. 10-11. Approximately fifty of the employees were holding jobs of each level, but the figure shows the percentage shift as higher score brackets are considered. For example, of those making test scores of 90 or better 27 per cent were holding Grade III jobs. However, if only those scoring 150 or better are considered, 52 per cent were holding Grade III jobs and the percentage continues to increase with higher critical scores until, when a minimum of 190 is considered, 85 per cent have attained Grade III jobs. Similar curves are presented for Grade II and higher jobs and for Grade II2 and higher jobs. Thus the curves can be interpreted as follows: If a critical score of 150 were imposed, within a five- to ten-year period, about 95 per cent would attain a Grade II or higher job, 85 per cent would attain a Grade II2 or III job, and 52 per cent would attain a Grade III job. The hiring implications where a promotional sequence is involved seem clear. Hiring specifications should be

established not only in the light of the demands of present jobs but also with a thought to the demands of the promotional sequence and procedure.

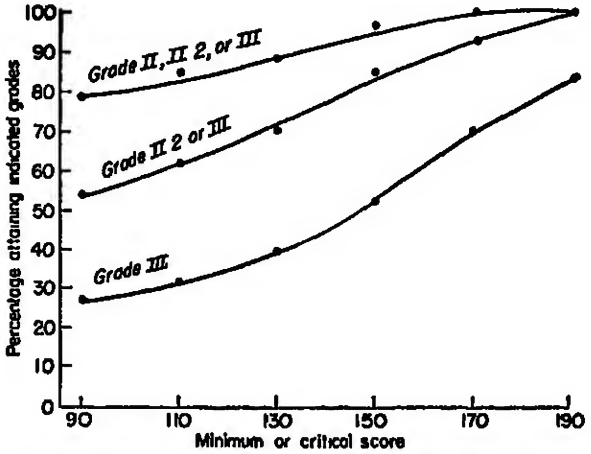


FIG. 10-11.—Proportion of clerical employees attaining or exceeding specific test scores who attained various clerical job grades in five to ten years on the job. (*From the Life Office Management Association.*)

SUMMARY

Clerical tests are among the most widely used and least frequently validated of all personnel tests. Generally, they are useful in identifying successful or potentially successful employees in four general classifications: clerks, typists, stenographers, and business-machine operators. Test validation in this area is just as essential as in any other; and although criteria are not easily found, some measure of job success is usually available. Evidence regarding the possibility of correlating job analysis and factor analysis in the clerical and office-work field has been presented.

CHAPTER XI

TESTS FOR SALESMEN AND RETAIL-STORE EMPLOYEES

In starting a discussion of tests for the selection of salesmen, it is appropriate to quote Shartle's statement[1] that "we have studied something like 20,000 occupations and we find that the difficulty encountered in devising improved selection techniques for sales work is probably not equaled in any other group of occupations." The differences among jobs all of which fall in the "selling" bracket are extensive, and many authors and experimenters including Rosenstein[2] have indicated the impracticality of trying to make generalizations drawn in one field of sales activity applicable to the selection of salesmen in another field. This, of course, is only a restatement of the whole point of view behind this book, namely, that validation for every job is a must.

INSURANCE SALESMEN

Life-insurance Salesmen.—More published work has appeared in the insurance field than in any other sales area, and the majority of this work has pertained to life-insurance salesmen.

Generally speaking, tests of the temperament and interest variety have been the most successful. The work of Strong[3] and others cited in Chap. VII is pertinent. Schultz[4] has used the temperament approach with some success as was indicated in

[1] SHARTLE, C. L. The measurement and selection of salesmen. *Mgmt. Rev.*, 1944, 33, 92–95.

[2] ROSENSTEIN, J. L. *The scientific selection of salesmen.* New York: McGraw-Hill, 1944. 162 pp.

[3] STRONG, EDWARD K., JR. *Vocational interests of men and women.* Stanford University, Calif.: Stanford University Press, 1943.

[4] SCHULTZ, RICHARD S. Standardized tests and statistical procedures in selection of life insurance sales personnel. *J. appl. Psychol.*, 1936, 20, 553–566.

Chap. VI. Using Beckman's *Revision of the A-S Test* and the Root's *Introversion-extroversion Test* he was able to discriminate with some degree of accuracy between the best and poorest producers among 259 life-insurance salesmen. Using as acceptable scores that range from the twentieth to ninetieth percentile on each test he obtained the results presented in Fig. 11-1. Of

FIG. 11-1.—Proportion of salesmen at three different production levels who made acceptable and unacceptable test scores. (*From Schultz.*)

those who were considered the best 20 per cent in production 68 per cent were in the acceptable score range, whereas of those who were considered the poorest 20 per cent in production 47 per cent had acceptable scores.

Steward's[1] system for selecting life-insurance salesmen has been widely recognized. His present system[2] for selecting life underwriters consists of "a composite inventory and examination." His statement of its contents follows:

1. Short form of the *Otis Self-administering Test of Mental Ability*
2. A modified form of Bernreuter's *Personality Inventory* to measure (*a*) dominance, agressiveness, and initiative and (*b*) stability
3. A general knowledge examination consisting of thirty-two items in the following fields: (*a*) economics and finance, (*b*) business arithmetic, (*c*) business law, and (*d*) home and social problems
4. A Vocational Interest in Selling Inventory

[1] STEWARD, VERNE. *The use and value of special tests in the selection of life underwriters.* 1116 East Eighth St., Los Angeles, Calif., 1934. 93 pp.

[2] STEWARD, VERNE. The development of a selection system for salesmen. *Personnel,* 1940, 17, 124–136.

5. A new Personality Trait Illustrations Inventory
6. A personal history section
7. A rating form

By weighting these various instruments, Steward has been able to predict sales production with a high degree of success. In a study [1] involving ten large metropolitan agencies he was able to demonstrate the relationship shown in Fig. 11-2. Those under-

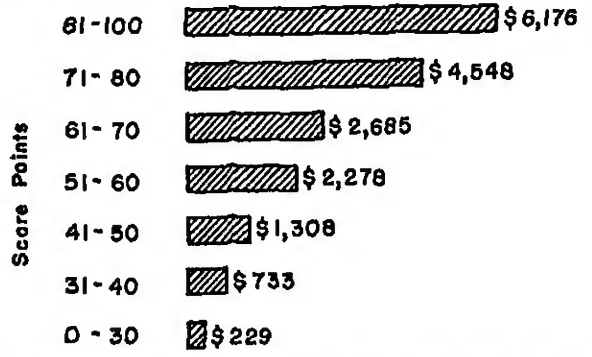


FIG. 11-2.—Average volume production of life underwriters scoring at various levels on Steward's system. (*From Steward.*)

writers who received between 81 and 100 points averaged sales totaling $6,176 for the period studied; those scoring between 71 and 80 sold an average of $4,548. There is a consistent decline in average sales with lower point scores, and those who score 30 or less average only $229 for the same period of time.

The work done by the Life Insurance Sales Research Bureau and reported by Kurtz [2] is perhaps the most comprehensive of all the work in this field. Extensive research, a discussion of which is beyond the scope of this treatment, has resulted in the development of the *Aptitude Index* to measure aptitude for life-insurance selling. The index consists of a personal-history rating form which incorporates some of the questions usually asked on an application blank and a personality-characteristics section

[1] STEWARD, VERNE. *Analysis of sales personnel problems.* Los Angeles: Verne Steward & Associates, 1943.

[2] KURTZ, ALBERT K. Recent research in the selection of life insurance salesmen. *J. appl. Psychol.*, 1941, 25, 11-17.

which is composed of a number of temperament and interest types of questions. The *Aptitude Index* for life-insurance salesmen has been validated against both sales records and agent-survival records, since many people who take up insurance selling are not sufficiently successful to remain in the business.

Figure 11-3 from the study by Kurtz [1] shows the production of 211 men who were twenty-six years of age or older. The figure is interpreted as follows: Those with A ratings who survived

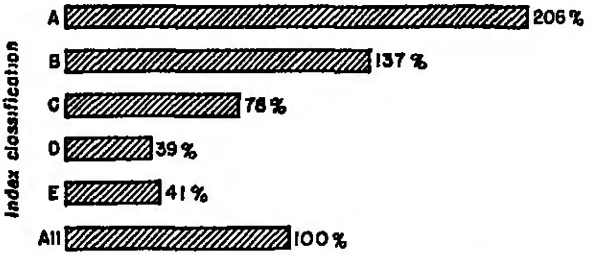


FIG. 11-3.—Percentage of average production of a group of 211 life-insurance salesmen classed according to *Aptitude Index* scores. (*From Kurtz.*)

sold 206 per cent as much insurance as a group selected at random from the group. Additional research with the *Aptitude Index* continues to show significant results. On company [2] has reported the facts presented in Fig. 11-4 which is based upon the records of 177 agents hired by one company over a one-year period. The length of the bars indicates the relative number of men in each rating class who are required to produce a given volume of sales. For example, according to the company's records, six A men will produce as much in a given time period as will eighty-four men with E ratings. This company no longer hires E men; and in 1942, 92 per cent of the men they hired had A or B ratings on the *Aptitude Index.*

Casualty-insurance Salesmen.—Bills' study,[3] the results of which are presented in Fig. 7-6 and discussed in Chap. VII, pre-

[1] *Ibid.*

[2] *The value and use of the aptitude index.* Hartford, Conn.: Life Insurance Sales Research Bureau, 1945.

[3] BILLS, MARION A. Relation of scores in Strong's interest analysis blanks to success in selling casualty insurance. *J. appl. Psychol.*, 1938, 22, 97–104.

sents evidence that a similar degree of success is possible in the casualty-insurance field. Ghiselli[1] in one study involving twenty-nine casualty-insurance salesmen used a combined rating and production index as a criterion and obtained correlations of 0.38 with the CPA score on the Strong *Vocational Inventory Interest Blank for Men* and 0.32 with the Pressey's *Senior Classification Test.* These and other studies suggest that predictions

FIG. 11-4.—Number of salesmen having each *Aptitude Index* grade that is required to produce a given volume of business. (*From Life Insurance Sales Research Bureau.*)

can be just as successful in the casualty field as in the life-insurance field.

MISCELLANEOUS SALESMEN

Laundry-supply Salesmen.—Otis[2] has reported the results of a study involving seventeen men engaged in selling soaps and alkalies to laundries. Using gross sales as his criterion, he obtained a correlation of 0.50 with combined life-insurance-salesmen scores and real-estate-salesmen scores on the Strong *Vocational Interest Blank for Men.* He further found correlations of 0.31 between personal data scores and selling cost and of 0.37 between personal data scores and gross sales.

Office Equipment.—Todd[3] reporting on two years of experience with an unidentified battery of tests for salesmen of office equipment presents the values shown in Fig. 11-5. Of those

[1] GHISELLI, EDWIN E. The use of the Strong vocational interest blank and the Pressey senior classification test in the selection of casualty insurance agents. *J. appl. Psychol*, 1942, 26, 798–799.

[2] OTIS, JAY L. Procedures for the selection of salesmen for a detergent company. *J. appl. Psychol.*, 1941, 25, 30–40.

[3] TODD, GEORGE L. Aptitude testing 600 salesmen. *Industr. Marketing*, September, 1944, 29, 32–33, 138, 140.

meeting a certain standard on the battery 35 per cent were considered good and 15 per cent failures; of those not meeting the test standard 8 per cent were good and 25 per cent were failures.

Machine Accounting Methods Salesmen.—The study of Ryan and Johnson[1] which was reported in Chap. VII is of interest here. Figure 7-7 shows significant results in the use of the Strong *Vocational Interest Blank for Men.*

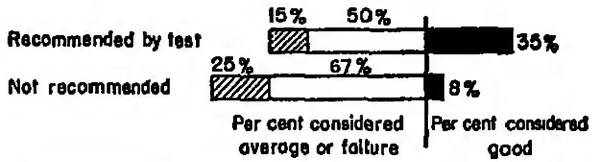


FIG. 11-5.—Proportion of office equipment salesmen classified on the basis of a test battery who were considered good (solid bars), average (open bars), and failures (shaded bars). (*From Todd.*)

Personality Tests.—The discussion of temperament tests in Chap. VI is pertinent with reference to salesmen. This is particularly true of Jurgensen's *Classification Inventory*[2] which is intended to be validated for each job and which promises to be quite useful with sales groups.

RETAIL-STORE PERSONNEL

Store Salespersons.—Using a modified scoring of "social dominance" on the Bernreuter *Personality Inventory,* Dodge[3] found that a critical score of 36 had significant discrimination value with store salespeople. A secondary group of male salespeople were rated as average, above average, and below average, and the proportion of each rated group falling above and below the critical score is shown in Fig. 11-6. The cut-off successfully identified 88 per cent of the above-average group and 75 per cent of the below-average group. Figure 11-7 shows similar results

[1] RYAN, T. A., AND JOHNSON, BEATRICE R. Interest scores in the selection of salesmen and servicemen: Occupational vs. ability-group scoring keys. *J. appl. Psychol.,* 1942, 26, 543–562.

[2] JURGENSEN, CLIFFORD E. Report on the classification inventory, a personality test for industrial use. *J. appl. Psychol.,* 1944, 28, 445–460.

[3] DODGE, ARTHUR F. Social dominance and sales personality. *J. appl. Psychol.,* 1938, 22, 132–139.

for a group of eighteen female salespeople. The results, though not so pronounced as those of the men, are nevertheless positive.

Store Cashiers.—This section dealing with cashiers and the following one, though not involving salespeople, are included in this chapter for want of a more logical position in the text. Clarke[1] using the *Otis Self-administering Test of Mental*

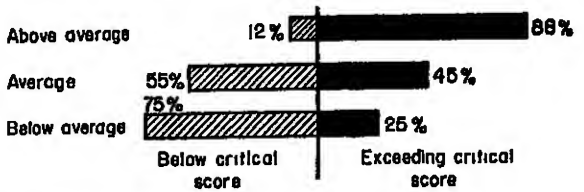


FIG. 11-6.—Proportion of rated ability groups of store salespeople (male) satisfying and failing to satisfy critical score on test (*From Dodge.*)

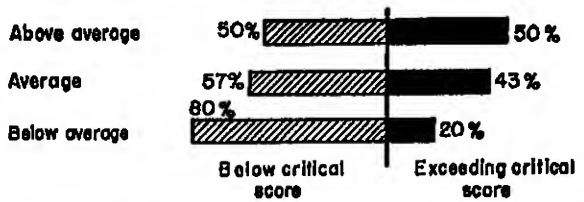


FIG. 11-7.—Proportion of rated ability groups of store salespeople (female) satisfying and failing to satisfy critical score on test. (*From Dodge.*)

Ability, a change-making test, and a dexterity test obtained a correlation of 0.59 between predicted production and actual production.

Viteles,[2] working with sixty-eight department-store cashiers and twenty-two inspector-wrappers, devised a test based upon job analysis. As described, the test included measures of the following: ability to follow directions, accuracy, arithmetic ability, common-sense judgment, and language ability. Viteles noted a distinct tendency for those who scored within the 30 to 70 bracket to have longer tenure. As shown in Fig. 11-8, nine people with scores less than 30 had an average tenure of 43 days;

[1] CLARKE, WALTER V. The evaluation of employment tests. *Personnel*, 1937, 14, 133–136.

[2] VITELES, MORRIS S. Selecting cashiers and predicting length of service. *J. Person. Res.*, 1924, 2, 467–473.

sixty-seven scoring in the 30 to 69 range had an average tenure of 140 days, and fourteen scoring 70 and higher had an average tenure of 61 days. Viteles further notes that "although the test distinguishes between competent and incompetent applicants, skill in handling the job does not increase proportionately with increased scores above the minimum passing score."

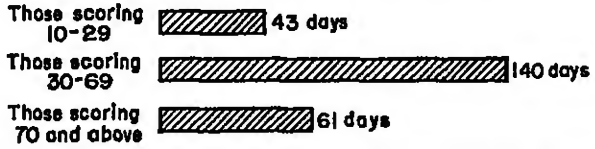


FIG. 11-8.—Average job tenure of cashiers and wrappers who scored at various levels on a test battery. (*From Viteles.*)

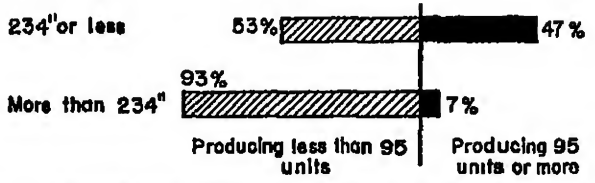


FIG. 11-9.—Relationship between test scores of department-store packers and production performance. (*From Blum and Candee.*)

Packers.—In a study involving sixty-five packers in a department store, Blum and Candee[1] used the *Minnesota Rate of Manipulation Test* against actual production scores and obtained the results presented in Fig. 11-9. Using 234 seconds as a critical score on the placing test, they demonstrated that of those performing better than that 47 per cent exceeded 95 production score units whereas of those doing poorer only 7 per cent exceeded this production standard. They further point out that this difference tends to disappear with experience.

SUMMARY

The fact that there appears to be no universal general sales type is a significant one in the personnel testing field. While it

[1] BLUM, MILTON L., AND CANDEE, BEATRICE. The selection of department store packers and wrappers with the aid of certain psychological tests. *J. appl. Psychol.*, 1941, 25, 76–85, 291–299.

is quite likely that there are job families among sales jobs as there are in other occupational areas, the practice of validating jobs on each sales situation is just as essential as is the application of the principle elsewhere. Mental ability tests may or may not be useful in a particular sales battery, and tests of the temperament and interest variety on the whole are the most useful.

CHAPTER XII

TESTS FOR SUPERVISORY, PROFESSIONAL, AND EXECUTIVE PERSONNEL

As the reader will readily recognize from the title, the present chapter represents something of a miscellaneous collection of jobs which it has not been possible to classify systematically in earlier chapters. Some professional jobs have something in common with some executive jobs, but the relationship between the others is not too apparent. The purpose of this chapter is to pull together some of the available studies treating occupations in these areas.

TESTS FOR SUPERVISORS

Mental Ability.—Every supervisor has things to learn; every supervisor has some "paper work" to do. Both of these facts suggest the importance of mental ability in the supervisor's job. Figure 5-13 showing the relationship between scores on the *Adaptability Test* and tenure supports this hypothesis. The fact that none of those men who scored less than 5 were on the job six months later and that 95 per cent of those scoring 15 or better were on this job is certainly significant. It must be kept in mind, however, that supervisory jobs vary tremendously. Some supervisors have more reports than others; some have a more extensive technology to learn than others; and some must utilize judgment and engage in creative thinking to a greater extent than others. All of these varying job requirements are no doubt related to the level of mental ability demanded by the supervisory job in question.

Harrell[1] has demonstrated the importance of general mental

[1] HARRELL, WILLARD. Testing the abilities of textile workers. Georgia School of Technology, *State Engineering Experiment Station Bull.*, July, 1940. 14 pp. Testing cotton mill supervisors. *J. appl. Psychol.*, 1940, 24, 31-35.

ability in supervision with a study of forty-two overseers in three different cotton mills. These men who had been rated as satisfactory or unsatisfactory by their supervisors were given the *Otis Self-administering Test of Mental Ability*. Test scores were converted to I.Q.'s by means of the standard practice, and Fig. 12-1 shows the results. When an I.Q. of 100 is considered as a critical score, 100 per cent of those meeting the critical score were among those considered successful whereas only 70 per cent

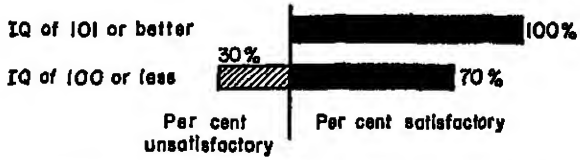


FIG. 12-1.—Proportion of supervisors above and below critical I Q. who were considered satisfactory and unsatisfactory (*From Harrell*)

of those failing to meet the critical score were considered satisfactory. This is ample evidence of the importance of mental ability in this particular job, but the fact that the prediction is as low as it is indicates that mental ability as measured by this test is by no means all that is needed in order to make a good supervisor. Although some studies similar to that reported by Sartain[1] have failed to show significant relationships between measures of supervisory success and selected tests, the facts seem to indicate that more often than not faulty criterion data or highly selected samples account for the findings. Shuman,[2] for example, has reported the successful use of the *Otis Self-administering Test of Mental Ability*, Bennett's *Mechanical Comprehension Test*, and the *Revised Minnesota Paper Formboard* for identifying those supervisors rated excellent in three different plants that are quite different in nature.

Attitudes and Interests.—Similar to the selling field, instruments that measure attitudes, beliefs, and interests seem most

[1] SARTAIN, A. Q. Relation between scores on certain standard tests and supervisory success in an aircraft factory. *J. appl. Psychol.*, 1946, 30, 328–332.

[2] SHUMAN, JOHN T. The value of aptitude tests for supervisory workers in the aircraft engine and propeller industries. *J. appl. Psychol.*, 1946, 30, 185–195.

promising at present for selecting supervisors. File's test[1] *How Supervise?* has already demonstrated its value in a few validity studies reported by File and Remmers.[2] One study pertained to forty-six successful supervisors and fifteen nonsupervisors who had been by-passed because of judged lack of ability. These men, employed by a company engaged in the manufacture of office machines, took one form of the test, and the results are shown in Fig. 12-2. This figure shows the percentage of each

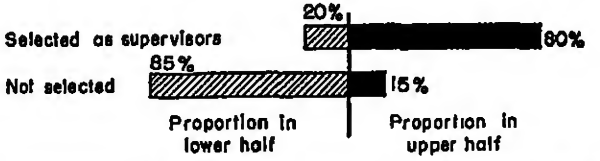


FIG. 12-2.—Proportion of men promoted and not promoted to supervisory position who scored in upper and lower half on *How Supervise?* test. (*From File and Remmers.*)

group that fell above and below the fiftieth percentile (median) of the published norms. In other words, of those selected as supervisors 80 per cent scored above average whereas of those judged not to have sufficient supervisory ability only 15 per cent scored above average on *How Supervise?*

In another study with the same test, twenty superior and twenty inferior supervisors were tested. The percentile equivalent of the mean of the better group was 96 and of the poorer was 63. Although both groups were above the average as represented by general norms, a significant difference exists between those individuals whom the company considers superior and those considered inferior.

PROFESSIONAL PERSONNEL

Engineers.—Studies dealing with the validation of test batteries for professional groups are rare, presumably for two

[1] FILE, QUENTIN W. The measurement of supervisory quality in industry. *J. appl. Psychol.*, 1945, 29, 323–327

[2] FILE, QUENTIN W., AND REMMERS, H. H. Studies in supervisory evaluation. *J. appl. Psychol.*, 1946, 30, 421–425.

reasons. (1) As a general condition, so few of any one professional group are employed with one company or their work is of such a diversified nature that validation in the accepted sense is

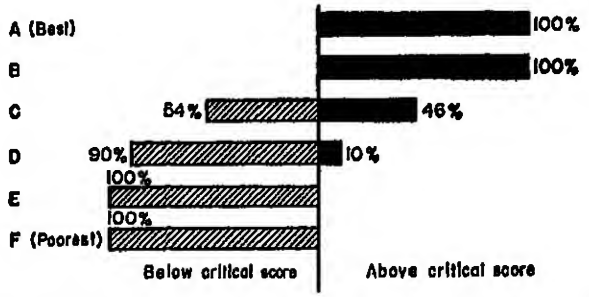


FIG. 12-3.—Proportion of engineers in each of six rated ability groups who were above and below critical score on vocabulary test. (*From Swartz and Schwab.*)

frequently next to impossible. (2) The second point hinges upon the fact that since all professions require extensive training, the training procedure has a tendency to eliminate by natural means the mentally and temperamentally unfit. It goes without saying, perhaps, that the more realistic and the more practical the training the more nearly true is this statement. One study reported by Swartz and Schwab [1] deals with a group of thirty-seven engineers who were rated by their supervisor on their ability as research engineers on a scale ranging from A to F. They were given the *Michigan Vocabulary Profile Test* and the *Minnesota Paper Formboard,* and the resulting relationships are shown in Figs. 12-3 and 12-4. Figure 12-3 shows the percentage in each rating group who were above and below a selected critical score on the three scientific vocabularies in the *Michigan Vocabulary Profile Test.* Note that all of the A and B rated people were above the critical score and that all of the E and F rated people were below the critical score. Figure 12-4 based upon the *Minnesota Paper Formboard* shows a similar trend which, though not so pronounced and regular, is nevertheless significant.

[1] SWARTZ, B. K., AND SCHWAB, R. E. *Experience with employment tests.* Studies in Personnel Policy No. 32, National Industrial Conference Board, Inc., March, 1941.

Nurses.—Dvorak's study [1] presents data on a superior and an inferior group of nurses. Figure 12-5 showing group profiles for superior and inferior nurses indicates significant differences on

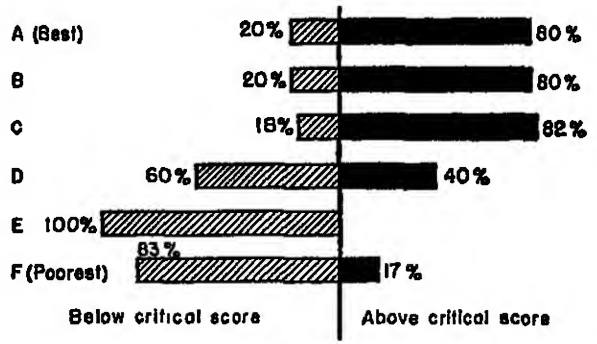


FIG. 12-4.—Proportion of engineers in each of six rated ability groups who were above and below critical score on *Minnesota Paper Formboard.* (*From Swartz and Schwab.*)

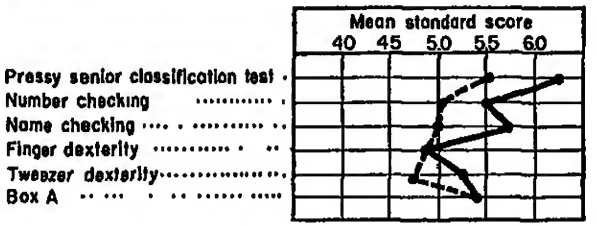


FIG. 12-5.—Group profiles on two ability groups of nurses. The solid line represents the test performance of the superior nurses; the broken line represents the inferior ones. (*From Dvorak*)

some of the tests. Two facts seem significant. (1) All but two of the tests seem to discriminate. (2) Where the tests do discriminate, the superior nurses are well above average.

EXECUTIVE PERSONNEL

The Executive Type.—According to Cleeton and Mason [2] there is no general "executive type" of personality. They draw

[1] DVORAK, BEATRICE. Differential occupational ability pattern. *Univ. Minn. Bull. Employt. Stab. Res. Inst.*, Vol. III, No. 8, University of Minnesota Press, 1934.

[2] CLEETON, GLEN V., AND MASON, CHARLES W. *Executive ability.* Yellow Springs, Ohio: The Antioch Press, 1946.

their conclusion from the fact that Strong and others have not been able to demonstrate a clear-cut pattern of interests and attitudes of persons in executive capacities vs. those in non-executive capacities. One of Strong's most recent studies,[1] though limited to public administrators, revealed extremely varied interests among the 517 men investigated. Cleeton and Mason do conclude that, in general, executives score relatively high on a wide variety of ability tests, thus indicating that individuals who successfully perform executive types of duties tend to be well-rounded personalities.

Probable Value of Tests.—A survey of the literature yields little evidence of successful validity studies in the executive brackets. This is no doubt due in part to the extreme difficulty attending the setting up of adequate criterion groups at the executive level, but it may also be partially attributable to a failure as yet to develop adequate measuring instruments. However, at present, in addition to mental ability tests, those which seem to be most promising are the temperament tests, the interest tests, and, to be specific, the *Michigan Vocabulary Profile Test.*

SUMMARY

In the selection of tests for supervisory personnel mental ability tests generally "come through." In supervisory, professional, and executive batteries, tests of the temperament and interest variety seem to be quite useful. It is probable that insufficient use has been made of vocabulary tests in executive and professional batteries.

[1] Strong, Edward K., Jr. Differences in interests among public administrators. *J. appl. Psychol.*, 1947, 31, 18–38.

CHAPTER XIII

HOW TO CONSTRUCT A TEST

The discussion thus far has dealt primarily with commercially available tests and their application in business and industry. However, individual plants, groups of plants, and trade associations are finding more and more need for tests constructed to meet their own needs. In addition, in some instances there is a tendency for organized labor to agree with management to effect certain promotions and transfers on the basis of test results.[1] Particularly in the trade test area, the need for tailor-made trade tests for the purpose is apparent. In addition, there is a growing tendency toward the selective scoring of commercially available tests for a given purpose. Particularly is this true of temperament tests in the selling area where it is less important to identify temperament components than it is to find those items and answers which actually discriminate between two criterion groups of salesmen. Whether it is a matter of test construction or the selection of significant items in a commercially available test, the problem of item validation is one and the same. The purpose of this chapter is to give some help in original item construction and to offer a simple but valuable method of item validation.

PREPARING A TEST BUDGET

Testing as Sampling.—All testing should be thought of as sampling. When one starts to test an individual's knowledge of a particular trade field, it naturally is quite impossible to ask all of the potential questions covering all of the possible elements of information. Not only would this be impractical; it would be impossible. An analogy of a housewife making cake icing is useful. After she has the icing made, she adds a few drops of vanilla

[1] See Chap. IX, pp. 127–128.

extract, stirs it well, then tastes it to see if she has added enough vanilla. She does not have to *eat all of the icing to tell.* She knows that one little taste gives a good indication of the whole batch *provided it is well stirred,* and the taste is consequently a representative sample. The same principle holds in testing. A few questions are set up by which we "taste" the knowledge of the applicant's or employee's knowledge. The estimate of his knowledge so gained is accurate insofar as the sample is representative. But again, like the cake icing analogy, if the sample is not representative, if the icing is not well mixed, spurious results may be obtained.

Content Outline.—In preparing a test over a particular trade content, then, the first step is to prepare an outline. Any trade can be broken down into a number of large areas or blocks. Each block, in turn, can be subdivided into minor units or areas. The following outline which was the basis for an auto-mechanic's test is an example.

OUTLINE OF AUTO MECHANIC'S TRADE

A. Power Plant
  1. Tune-up
  2. Cooling system
  3. Fuel distribution
  4. Engine rebuilding
B. Electrical System
  1. Basic information
  2. Batteries
  3. Starting motor and generator
C. Front End and Steering Mechanism
  1. Steering gear
  2. Suspension system
  3. Applied geometry
  4. Service
  5. Ride control
  6. Bearings
D. Chassis
  1. Transmission
  2. Clutch
  3. Universal joints
  4. Rear axle
  5. Brakes
  6. Lubrication

Jobs of lower than trade level can similarly be broken down. For example, one airframe plant broke riveting down into four major headings: drilling, riveting, subassembly, and skin riveting. Each of these in turn was broken down into subparts. Outlines such as these are essential if the sampling principle is to be followed. If an outline is not made, some areas are almost sure to be omitted and others are apt to be overemphasized. The importance of tradesmen and trade advisory committees in preparing such an outline cannot be stressed too much. Frequently union participation at this stage ensures union acceptance later.

Further Analysis.—Although the content outline is essential, it alone does not ensure the preparation of the right questions or items. Referring to the above outline, the question may be raised "What kinds of items can be prepared on tune-up?" The following outline adopted from one developed for preparing trade analyses for deriving training content is useful:

WHAT A WORKER NEEDS TO KNOW

A. Work Procedures
  1. The "how to do it" question
  2. The "which should be done first" question
B. Technical Skills
  1. Reading drawings
  2. Making sketches
  3. Making drawings
  4. Laying out patterns
  5. Making calculations
C. Job Information
  1. Trade terms
    a. Names of material and stock
    b. Operating terms
    c. Location names
    d. Machine and machine-part names
  2. Care of tools
    a. In use
    b. Not in use
    c. Prevention of loss
  3. Information about stock
    a. Recognition properties
    b. Working properties
  4. Job science
    a. Mechanics
    b. Hydraulics

c. Pneumatics
d. Electricity
e. Heat
f. Light
g. Strength of materials
h. Chemistry

5. Safety information
   a. Prevention of injury to self
   b. Prevention of injury to others
   c. Prevention of injury to equipment

Obviously, no single outline can be prepared that will serve equally well all job and occupational areas. However, some breakdown of what the worker needs to know similar to the one above makes the preparation of questions easier.

The Test Budget.—In preparing the test the next step is making the test budget. This budget may take the form of a chart, similar to the one in Fig. 13-1, which actually combines the two outlines previously given. A chart of this character can then be completed by entering in each cell the proposed number of items. This phase can best be handled through some kind of a group or committee approach. Some companies are finding labor-management committees desirable. The one important point is to be certain that the committee includes representatives who are competent in the job area in question. It must be remembered that there is no magic substitute for judgment. The only way that values can be arrived at for inclusion in this outline is through judgment.

CONSTRUCTING ITEMS

Types of Questions.—Up to this point the discussion has dealt solely with the subject matter to be tested. Now comes the question of what type of item to use. The unreliability of the so-called "essay question" is well known and will not be discussed here. Some form of objective, single-answer question is, of course, necessary. There are many such types, and they have been thoroughly discussed and illustrated by Remmers and Gage.[1] This discussion will be confined to those types of items

[1] REMMERS, H. H., AND GAGE, N. L. *Educational measurement and evaluation.* New York: Harper, 1943.

which have been most useful in industrial tests of the trade type. These are the true-false or yes-no item and the multiple-choice

	ITEM BUDGET FOR AUTO MECHANICS TEST	WORK PROCEDURE	INFORMATION					SAFETY INFORMATION	TOTAL
			TECHNICAL SKILLS	TRADE TERMS	CARE OF TOOLS	STOCK INFORMATION	JOB SCIENCE		
	Tune-up								
POWER PLANT	Cooling system								
POWER PLANT	Fuel distribution								
POWER PLANT	Engine rebuilding								
ELECTRICITY	Basic information								
ELECTRICITY	Batteries								
ELECTRICITY	Starting motor and generator								
FRONT END AND STEERING	Steering gear								
FRONT END AND STEERING	Suspension system								
FRONT END AND STEERING	Applied geometry								
FRONT END AND STEERING	Service								
FRONT END AND STEERING	Ride control								
FRONT END AND STEERING	Bearings								
CHASSIS	Transmission								
CHASSIS	Clutch								
CHASSIS	Universal joint								
CHASSIS	Rear axle								
CHASSIS	Brakes								
CHASSIS	Lubrication								

FIG. 13-1.—Blank form for use in preparing test item budget.

item. Others can and have been used, but these two are generally speaking most useful, and discussion will be limited to them. The true-false type is illustrated by the following:

	True	False
1. A storage battery has separators between the plates - -	☐	☐

The testee is asked to indicate whether the statement is true or false by making an *X* in one of the squares. An alternative is to ask a question that can be answered with yes or no. The major advantages of this type of item are its simplicity, its ease of construction, and the fact that more of this type can be answered in a given time period than is the case with any other type.

The multiple-choice or multiple-response item is illustrated by the following:

1. In the formula, $4d=7d+15$, $d=$ - - - - - - - - - (___)
   (1) +5, (2) −5, (3) +11, (4) −11

The testee merely enters 1, 2, 3, or 4 in the blank to indicate the answer that he believes is correct. Sometimes four alternatives are used; sometimes three; and sometimes five. All facts considered, four seems to be the optimum number. Although it is more difficult to construct than the true-false item, it is the best all-around type of item because it is least subject to guessing and consequently is more stable. Whereas with a true-false item the chances of guessing are fifty in a hundred, with the four alternate multiple choice item the chances are reduced to twenty-five in a hundred. As indicated before, there are instances where other types of items are as good or even better, but generally speaking it is the most useful type of item for personnel tests.

Suggestions for Preparing Items.—Numerous books on psychological measurement are available, many of which provide usable suggestions. Notable among these is the one by Remmers and Gage [1] already cited. A few of the more significant points with respect to multiple-choice item construction seem to be in order, however.

1. Prepare some hard, some easy, and some moderately difficult items.
2. Avoid trick or catch items.
3. Avoid obviously wrong alternatives.
4. Make each item as short as possible.
5. Make all alternatives or responses about the same length.
6. Place correct answers in random order.
7. Avoid "cue" words in the root of the item.

[1] *Ibid.*

Number of Items.—In the event that time and conditions permit, more items should be prepared than are desired in the test after preliminary tryout. Two to one is a good ratio. Thus, if a finished test of 100 items is desired, the experimental form should include 200, and the item budget similar to the one in Fig. 13-1 should provide for that number. This will permit the elimination of the less valuable and the bad items in accordance with the item-analysis procedure described below.

ITEM-ANALYSIS PROCEDURES

Item Difficulty.—Item-analysis techniques generally have two purposes: to determine the difficulty of the item or to determine the validity of the item. Item difficulty is measured in terms of the proportion of the population tested that gets the item right. Thus, if one item is passed by 93 per cent of the population while another is passed by 78 per cent of the population, the former is easier than the latter because the probability that any single person will get it right is greater. Thus, by simply counting the number of persons in an experimental population who get each item correct and converting these raw numbers to percentages, all items can be ranked in the order of difficulty. It is generally considered good test procedure, when a test is revised, to place the easiest item first, the hardest one last, and the intermediate ones in their corresponding positions according to percentage passing.

Item Validation.—Difficulty alone, however, does not determine the usefulness of a particular item. It is necessary to know whether or not the item is valid. Does it measure what the test as a whole measures, or does it measure what the test is supposed to measure? To evaluate the extent to which test items meet one or both of these criteria, several methods have been devised. An extensive discussion of these methods is beyond the scope of this chapter. However, one method of item validation that has been shown to be satisfactory will be discussed.

The Principle of Item Validation.—In evaluating the validity of test questions, a criterion is necessary. In other words, it is necessary to identify two groups of people, known or assumed to differ on the trait, attribute, or knowledge being measured, and

to compare the performance of the two groups on the item in question. If the good or superior group gets the item correct with greater frequency than does the poor or inferior group, the item is assumed to be good. However, if equal or nearly equal proportions of the two groups get the item correct or if, perchance, the poor group does better, it is assumed that the item is invalid for that purpose. Two methods are generally employed for selecting the criterion groups. These are known as the criterion of internal consistency and the use of an external criterion.

Use of External Criterion.—When an external criterion is used, some means, in no way related to the test, is found for identifying two groups that are known to be different and the performance of each group on each item in the test is determined. For purposes of illustration, the performance of naval trainees on the *Practical Electrical Information Test* discussed in Chap. IV will be used. An experimental form of the test, consisting of 116 items, was administered to 237 trainees at the start of the training program. Four sample true-false items with their original numbers follow:

156. A return-call push-button system requires three wires.
160. In rigid iron conduit, one side of the line is grounded.
195. A household electric meter records the voltage used.
228. When an automobile horn is used, less current goes through the button circuit than through the horn.

Items 156, 195, and 228 will be discussed first. Now these three items were equal or nearly equal in difficulty, since 59 per cent of the total group got number 156 correct and 53 per cent of the group got numbers 195 and 228 correct. However, as determined later, they differ considerably in validity. After the fifteen weeks' training period had passed, and after each of the 237 men had made a record in the school, the group was divided into the 25 per cent making the highest grades in the school, the 25 per cent making the lowest grades, and the middle 50 per cent. A study was then made of the proportion of each grade group who got each item correct. Figure 13-2 shows the results. Note that on Item 195, 36 per cent of the low group, 54 per cent of the middle group, and 77 per cent of the high group answered the

item correctly. This is then a valid item. Note also that a similar relationship exists on Item 156 but that the spread is not so great. This is still a valid item, but it is not so good as 195. Note, then, the results on number 228. The percentages of 53, 54, and 51 indicate that the performance of the three groups is essentially identical. In other words, a member of the superior

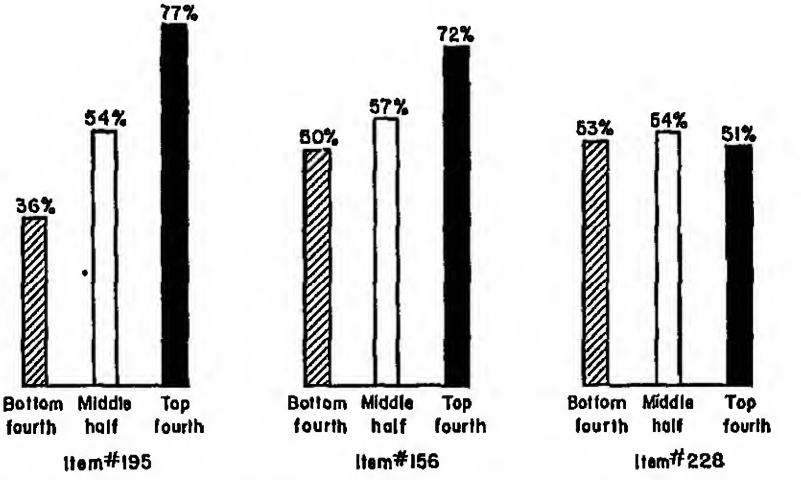


FIG. 13-2.—Percentage of trainees receiving course grades in the bottom fourth, middle half, and top fourth of the class who passed each of three questions in an experimental form of the test.

group is no more likely to pass the item than is a member of either of the other groups. This item is said to be invalid and is discarded.

Figure 13-3, based upon the results obtained for Item 160, illustrates the type of reversals that may occur. In this instance the poor group actually surpassed the good group, since 73 per cent of the former and 61 per cent of the latter got the item right. Obviously, items of this sort operate at cross purposes with the rest of the test and must be eliminated for best results.

Discrimination Values.—While the items used above for illustrative purposes could be accepted or rejected by inspection, it has been found helpful to use an index to describe the validity of

test items. Figure 13-4 is a nomograph [1] designed to assign such an index known as the *discrimination value*, or *D*-value. While Figs. 13-2 and 13-3 show the performance of three groups, high, low, and middle, for purposes of determining a *D*-value only the two extreme groups are necessary. The nomograph is used as follows: (1) Look up the percentage of the high group passing

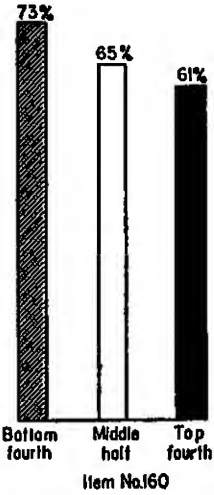


FIG. 13-3.—Percentage of trainees receiving course grades in the bottom fourth, middle half, top fourth of the class who passed a given question in an experimental form of the test.

on the scale at the left and the corresponding value for the other group on the scale at the right; (2) connect the two points with a straightedge; (3) read the value on the middle scale where the straightedge intersects the scale line. Thus, the *D*-values for the items used earlier are: 195, 1.1; 156, 0.6; 228, 0; 160, −0.3. Note that in the case of Item 160 the *D*-value is negative; negative signs always appear when there is a reversal, that is, when the poor group does better than the good group. In this fashion every item may be assigned a *D*-value that is indicative of its validity. As will be shown later, it is then a simple matter to select the most valid items for inclusion in the revised form of the test.

[1] LAWSHE, C. H., JR. A nomograph for estimating the validity of test items. *J. appl. Psychol.*, 1942, 26, 846–849.

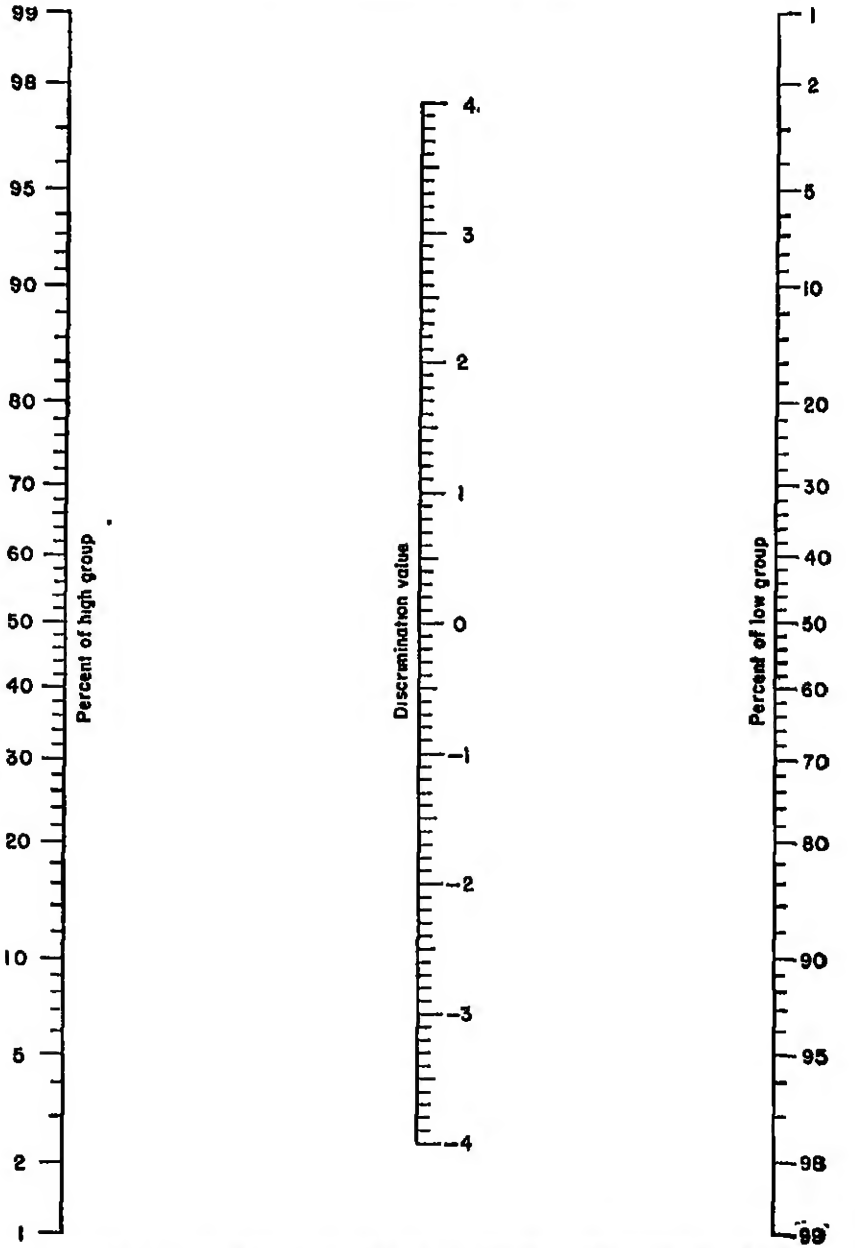


FIG. 13-4.—A nomograph for assigning *D*-values (discrimination values) to test questions. (*From Lawshe.*)

Criterion of Internal Consistency.—Even though a trial or experimental form of a test is almost certain to include some poor items, the best assumption is still that those getting the highest scores are the best and those getting the lowest scores are the poorest. For this reason extreme scoring groups are frequently taken as criterion groups. For example, if 200 people take a test, even though there are certainly some bad items, there is still little question that the 50 scoring highest are superior in whatever the test measures than are the 50 scoring lowest. Consequently, it is sometimes customary to set up the high 25 per cent and the low 25 per cent, the high 30 per cent and the low 30 per cent, or other extreme proportions as criterion groups with reasonable assurance that the two differ in the quality or knowledge being measured. The problem then becomes one of comparing the performance of these two groups on each item of the test.

For purposes of illustration, two items from an experimental form of an interest inventory for sales personnel are presented. This inventory was administered to 300 candidates for sales jobs of varying types. The items were scored on the basis of the best logical analysis that could be made, and the 100 persons making the highest scores and the 100 making the lowest scores were identified and designated as criterion groups. One question was as follows: "Do you like to think a long time before expressing your opinions?" Provision was made to answer "yes," "no," or "maybe." The "no" answer was given by 67 per cent of the high group and by only 22 per cent of the low group. This item, as reference to Fig. 13-4 will show, has a *D*-value of 1.2 and is highly valid. On the other hand, to another question that read "Do you rather enjoy spending an evening alone?" 42 per cent of each group answered "No." The *D*-value is obviously zero, and the item in no way discriminates between the two groups.

USING ITEM-ANALYSIS RESULTS

Revising the Test.—Both of the procedures outlined above are useful in identifying the most valid items. As pointed out before, the methods differ only in the means whereby extreme

groups of people are identified. Once *D*-values have been obtained, a distribution similar to the one in Table VII should be prepared. This is a frequency tabulation of the actual *D*-values obtained in the administration of the experimental form of the

TABLE VII.—FREQUENCY DISTRIBUTION OF *D*-VALUES OF ITEMS IN EXPERIMENTAL TEST

D-value	Tabulation	*f*
1 1	/	1
1.0	/	1
0.9	/	1
0.8	𝍸 //	7
0.7	𝍸 ///	8
0.6	𝍸 𝍸 ///	13
0.5	𝍸 𝍸 ///	13
0.4	𝍸 𝍸 //	12
0.3	𝍸 𝍸 𝍸 //	17
0.2	𝍸 ////	9
0.1	𝍸 𝍸 //	12
0.0	𝍸 𝍸 ////	14
−0.1	////	4
−0 2	/	1
−0.3	//	2
−0.4		0
−0.5		0
−0.6		0
−0.7		0
−0.8	/	1
		Total....116

Practical Electrical Information Test described above. Notice that the *D*-values range from 1.1 to —0.8. The next problem of the test maker is to decide how many items to retain. This is a question for which there is no generally applicable answer. Generally speaking, a long test is better than a short test but, of course, not if the long test is simply the short one with poor items added. The problem then becomes one of starting with the highest *D*-value item and moving down until enough items have been included to make a test of reasonable size. In the example at hand, a cutoff at 0.4 was used. There were fifty-six items with

D-values of 0.4 or better. One of these was discarded for a reason not associated with the present discussion, leaving fifty-five to be included in the revised test. As indicated in Chap. IV, this short test proved quite useful in identifying potentially successful naval electrical trainees. Arbitrary cutoffs are hazardous because of the differences in pairs of criterion groups. However, experience has indicated that in the average situation, *D*-values

TABLE VIII.—DISTRIBUTION OF TIME REQUIRED BY 237 PERSONS TO COMPLETE A TEST

Min.	Tabulation	*f*	*cf*	Percentile
15	/	1	237	100
14		0	236	99
13	𝍸 𝍸 ////	14	236	99
12	𝍸 𝍸 𝍸 𝍸 𝍸 /	26	222	94
11	𝍸 𝍸 𝍸 𝍸 ///	23	196	83
10	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	40	173	73
9	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 //	47	133	56
8	𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸 𝍸	45	86	36
7	𝍸 𝍸 𝍸 𝍸 𝍸	25	46	17
6	𝍸 𝍸 /	11	16	7
5	////	4	5	2
4	/	1	1	0

less than 0.3 or 0.4 seldom contribute to a more reliable or valid test. The problem still remains one of balancing maximum-size *D*-values with maximum test length.

Establishing Time Limits for Tests.—As a general principle, mental ability tests are the only ones in which speed is a factor. Consequently, the reader will rarely if ever need to consider speed in score determination. However, even on a straight-forward trade information test, some time limit is necessary. It is customary to set such time limits so that there is little emphasis on speed and so that nearly everyone can finish. This can be achieved by administering the test on a trial basis with no time limit and asking each person taking the test to record the time when he finishes. These time values together with the starting time can

be used to tabulate the time required for the total group. Table VIII is a sample distribution. Note that the time required ranged from four to fifteen minutes. By means of the tabulations and the values in the accompanying columns, a reasonable time can be established. Since only 73 per cent have finished in ten minutes, that is obviously not enough time. In this particular test thirteen minutes was established. Note that 99 per cent (all but one of 237) have finished. To extend the time two more minutes for one person seems unreasonable. When knowledge or information types of tests are being used, as indicated earlier, items are usually arranged in order of difficulty with the most difficult items at the end. Since there is some relationship in this kind of test between rate and accuracy, the probability that these very slow people will get the last items correct is small. Therefore as a general principle no damage is done by setting the time limit so that even as many as the slowest 5 per cent cannot finish.

CHAPTER XIV

INAUGURATING AND OPERATING A TESTING PROGRAM

The preceding thirteen chapters in this book have dealt with the technology of personnel testing. Every personnel man and every other person who has held a responsible job in business or industry know that good technology is frequently scuttled because of improper methods of initiation and operation. The purpose of this final chapter is to crystallize a workable point of view for test program administration.

First Considerations.—*No company should undertake the use of psychological tests without the assistance of a person specifically trained in the use of such tests.* Generally speaking, the management has three alternatives: (1) to add a well-rounded, fully qualified personnel psychologist to the staff; (2) to secure the assistance of a qualified consultant[1] who is on the staff of a university or a responsible consulting firm; or (3) to select a member of the present personnel organization who has some interest or background in the field and send him to a recognized university for training. Frequently a combination of the last two is the best solution if the company feels that it is not large enough to justify the addition of the services of a full-time psychologist.

Too often companies have "dabbled" with tests when no individual in the organization was qualified beyond being "interested." Instances of this character more often than not lead to trouble and frequently result in tests being "kicked out." A

[1] While there are many qualified persons doing consulting work in the testing field, there are also many so-called "consultants" who have neither the training nor experience to do the job. Reputable psychologists are usually members of the American Psychological Association, 1515 Massachusetts Ave., N.W., Washington 5, D.C.

testing program built on solid ground and guided by a competently trained individual never experiences this misfortune and usually grows in importance and respect.

THE BASIC PROCEDURE

At the risk of oversimplifying the problem, a basic procedure consisting of seven fundamental steps will be outlined here. They are

Step 1. Establish the personnel-testing policy

Step 2. Introduce the program in the plant

Step 3. Identify the jobs or departments having personnel problems

Step 4. Obtain job descriptions for these jobs

Step 5. Select tests for tryout

Step 6. Select criterion groups and administer tests

Step 7. Establish the operating test program

Each of these will be discussed individually.

Step 1. *Establish the Personnel-testing Policy*

Management Support Necessary.—Few if any testing programs succeed without the support of top management, and few that have the support of top management fail providing technical competence is available. Authority, however, should not be confused with support. Many managements have authorized programs of one kind or another without whole-hearted support. The result is frequently failure. Top management must genuinely believe that personnel tests are tools that are essential in the administration of an enlightened personnel policy.

Getting Management Support.—How can management be convinced that tests are deserving of managerial support? In

general, this support can be obtained in one of two ways: through the experience of other companies or through pilot studies in the plant in question. Frequently this latter approach is extremely effective. If one department or one operation where a severe personnel bottleneck exists can be selected for a tryout and real results can be shown, almost any management will sit up and take notice. Evidence of this kind supported by forecasts of future employment and personnel problems confronting the company in which tests might help is useful.

Budget, Location, and Policy.—In any far-reaching program a budget is, of course, necessary. The failure of management to establish a budget, however modest, is evidence of lack of the support mentioned earlier. Where the testing program should be located and what department or administrative officer should be responsible are questions that cannot be answered without a knowledge of the variables involved. Obviously the function should be located somewhere in the industrial-relations department or division. Whether it is attached to personnel, training, or employment or is a separate unit must be decided in terms of company size, qualifications of personnel in these departments, and the particular emphasis that the company wishes to place on testing. Wherever the placement, the beginning should be modest. Physical facilities and personnel should be geared to present needs and should not be overambitious; when testing produces results, growth will come naturally.

Written Policy Desirable.—Policies are always better when they are written, and personnel-testing policies are no exception. Once management has decided to support a testing program, the purpose of the testing program, what test data are to be used for, and the responsibilities of those in charge should be made a matter of record. Below is a sample policy for a company sufficiently large to support a separate testing unit known in this company as a division.

GENERAL POLICY OF DIVISION OF PERSONNEL TESTING

The purpose of this policy is to define the function of the Division of Personnel Testing in regard to its furnishing information

and recommendations on the selection and placement of personnel and the establishment of personnel standards for given occupations and jobs.

I

The Division of Personnel Testing is authorized to and is responsible for the securing of objective information from applicants for employment and from general employees as directed by the personnel manager. The function of the division is one of gathering information with the use of objective and scientific methods and reporting on it. Its function is *not* one of making decisions.

II

As a consulting service, the Division of Personnel Testing may assist any authorized department or supervisor in securing and analyzing information on any applicant, employee, or group of employees for the purpose of determining qualifications for employment, transfer, and assignment and for setting job standards for personnel. Such information will be so presented that an objective and complete appraisal can be arrived at without prejudice or harm to the character or reputation of the applicant or the employee.

III

The personnel manager shall be responsible for the establishment and maintenance of a committee of responsible representatives from operating supervision and the personnel department. This committee on personnel testing shall be responsible for the establishment of a program to meet the needs of the company for this service and shall be responsible for the program's being carried out efficiently and effectively.

IV

The Division of Personnel Testing will keep accurate and dated records of all its testing and recommendations. The division shall keep all of its records on persons tested strictly confidential, available only to authorized supervision and manage-

ment. It shall use only methods and materials acceptable to its profession and shall conduct its affairs and make its reports on a strictly objective basis. It shall not discriminate against or be prejudiced by any person, applicant or employee, because of race, creed, color, country of origin, or union affiliation.

V

Any supervisor or other person in authority shall feel free to discuss, in line with his work and his own subordinates, the test results and their meaning with any of his people. The function of the division is to make available such information that it might have or be able to secure and to interpret it upon proper request. The division will not give out abstract test results as such, nor will it make any decision with respect to any individual tested. It is, in every instance, the responsibility of the supervisor or the committee on personnel testing to make decisions concerning the hiring, placement, or movement of employees.

VI

Any employee of the company should feel free to discuss his own test results with the supervisor of personnel testing if he has a problem that he feels will be benefited by such a conference. No test results will be given to the employee; their meaning and significance will be discussed only according to acceptable counseling procedure. No test results will be available to any outside agency without the written consent of the employee concerned. Test results of applicants are not available to the individual. No test material may leave the division or its jurisdiction.

VII

No provisions of this policy on personnel testing shall be changed or added to without the written approval of the plant manager after such recommendation for change has been made in writing by the committee on personnel testing. A budget shall be set up within the personnel department for the Division of Personnel Testing, and the division will stay within its budget.

The exact nature of the policy will vary from plant to plant; and although the example above would not be satisfactory in all situations, it nevertheless illustrates the point. Such a policy should be clear, direct, and brief.

Step 2. *Introduce the Program in the Plant*

Function of Tests.—The reader should continually keep in mind the fact that tests are only a tool in the employment situation. The facts that they supply are only additional information which, together with application and interview data, contributes to better employee selection and placement. It is important that test data (and data from other sources) should not be used to force a particular employee down the throat of the supervisor. The most successful approach is to (1) screen out obviously unfit applicants, (2) identify those most likely to succeed on a particular job by means of test and other data, and (3) refer the candidate or candidates, together with data, to the supervisor in question. *The supervisor should make the final decision.* In the last analysis the supervisor must live every day with the choice, and he alone is responsible for success or failure in his department. To take away from him his right to choose those who are to report to him is to deprive him of one of his important supervisory functions. *Tests are only aids and should not in any way be used to shear the supervisor of any of his authority.*

How to Get Supervisory Support.—In general, the same techniques are effective in getting the support of supervisors as are useful in getting the support of top management. Top management authority is no substitute for supervisory support. Supervisors in any plant in which testing is being inaugurated will vary in their attitudes all the way from violent rejection to enthusiastic acceptance. One of the first jobs of the testing-program director is to tone down the enthusiasm of the latter and "sell" the former on the value of tests. This can best be done through pilot programs. That is, a situation should be selected where there is a real personnel problem and in which the particular supervisor is willing to "go along." Once results have been obtained, they should be presented to other supervisors by

means of the kinds of graph and chart used in this book. Slowly the program will grow, and eventually one supervisor after another will request testing services. There is little or no merit in swinging an entire plant into a testing program all at once. The success of a testing program is not measured in terms of such statements as "We tested *so many* people last month." Success must be evaluated in terms of the number of job classifications for which batteries have been validated plus the genuineness with which the program is accepted by supervisors and others.

How Not to Evaluate Tests.—Although there is usually no harm in permitting supervisors to look over or even take tests, the person in charge of testing should not get himself in the position of letting supervisors decide on the merits of a test in terms of how it looks. Supervisors should be trained to accept or reject tests on the basis of facts similar to those presented in this book and not in terms of whether or not he likes a particular test or a particular question in that test.

Arguments with those who object to testing never accomplish anything of a constructive nature. If a particular supervisor cannot be won over on the basis of personal relationships plus facts, he certainly cannot be won by argument. Statements such as "Skeptics are the best friends of testing" and "If everyone was as skeptical as you, there would be fewer 'quacks' in the testing field" are often entrees to a satisfactory relationship.

Preemployment Testing and Organized Labor.—Preemployment testing is completely outside the jurisdiction of the union except in instances in which the closed shop functions and the union actually supplies employees to the company. When an open or union shop exists or when such union security devices as maintenance of membership are operating, management may select anyone whom it chooses and it may ask any questions that it wishes orally in the interview or in writing in the form of application blanks or tests.

Union Support Needed.—However, union support is necessary if the present employee method of test validation described in Chap. II is to be used. That is, if the company wishes to pull off

the floor a number of machine operators, presently on the job, for the purpose of setting test standards to be used in new hiring, union support is necessary.

Upgrading.—The use of tests in the upgrading or transferring of present employees is a matter of concern to union officials. It is a matter that can diplomatically be worked out as a general rule provided there is union representation in the early planning. However, if a plan is devised and union officials are asked to accept it, differences usually arise and sometimes testing is completely blocked.

Some unions, although not many, are on record as being opposed to tests as well as to other devices of scientific management. The vast majority, however, are open-minded. Although essentially all unions have gone "all out" for seniority as the criterion for promotion, a significant fraction of all contracts are written with a clause such as "where ability and physical fitness are relatively equal, seniority shall be the determining factor." Management at times has been weak in its ability to demonstrate the superiority of an employee. There are extremely promising signs throughout the country that many unions are becoming more willing to measure certain kinds of ability by means of tests.[1]

Most Unions Can Be Sold.—Where management is sincere, where there is a history of genuine collective bargaining, and where there is a reputation of managerial fair play, organized labor nine times out of ten will go along with a well-grounded testing program. If there has been excessive friction, tests will be dubiously received if at all. If management has a reputation (based upon real or imaginary happenings) of shady practices, testing will be looked upon with suspicion.

Step 3. Identify the Jobs or Departments Having Personnel Problems

Criteria for Choosing.—Tests are of no value if they do not help management to do a better job of managing. Unless a managerial problem can be solved or can be handled better as the

[1] See Chap. IX, pp. 127–128.

result of a testing program, pragmatically there is absolutely no point to testing. This statement, of course, assumes management's ability to identify problems. The testing program should be inaugurated in a modest way. One or two jobs or departments in which there are known problems should be identified. These should be selected because they are known to be high in turnover, high in training costs, high in absenteeism, high in wastage, or low in production or *for some other tangible reason.* Vague generalizations to the effect that "morale is low on that job" are of little help. Unless the problem is tangible, the approach to its solution tends to evade objectivity. Reasons for selecting a few problem spots of this sort are evident: (1) They are known to be problem spots throughout the plant generally, and (2) any improvement that is brought about is readily apparent.

Step 4. Obtain Job Descriptions for These Jobs

Job Analysis.—Job study after the fashion suggested in Chap. II is necessary at this point. It is highly important to take advantage of any prior job study that may already have been conducted in the organization. It is well to contact those in charge of methods and procedures, motion and time study, job evaluation, and safety because in many instances valuable job analyses have already been conducted. Rarely are these completely adequate, but quite often they provide a firm basis upon which to build, and it is not uncommon to find that the key factors in a job have already been identified.

Step 5. Select Tests for Tryout

Also as outlined in Chap. II, at this point several tests should be selected for tryout. The number and character of tests should be determined by the job itself and the relative importance of improved selection in the job.

Step 6. Select Criterion Groups and Administer Tests

Procedure.—The whole problem of selecting criterion groups has been extensively discussed in Chap. II, and little can be

added here. Extreme care should be taken in the case of the present employee method to make certain that everyone, the supervisor, the union official, and the individual employee, knows why the employees are being tested. It is well to explain again that "we are *testing the tests*" and that how the employee does will in no way affect his job.

Step 7. Establish the Operating Test Program

Evaluating Tests.—At this point, those tests which show an actual relationship with one or more criteria should be identified. Too much emphasis cannot be placed on the necessity of using one of those techniques (or a similar one) outlined in Appendix A for determining the extent to which the results obtained might be attributable to chance alone. The possibility of tests "backfiring" can be virtually eliminated if these checks are always applied.

Critical Scores.—In some instances it will be desirable to establish critical scores, that is, cutoff points designating those portions of the score range which are acceptable and those which are unacceptable. More often, however, because of fluctuating personnel demands and fluctuating labor markets, a more flexible scheme will be desirable such as A, acceptable; B, acceptable only under certain circumstances; and C, unacceptable.

Personnel Records.—The necessity of adequate personnel records is apparent. Those in charge of personnel research should set up some sort of summary sheet to be filed with the personnel record. In addition to test data, such a sheet should include transfer, promotion, rate change, and termination data. Absentee facts, a record of accidents, merit-rating results, and any other data likely to be useful for criterion purposes should be included. Such a sheet, filed with the personnel record, is conveniently located and yet may be withdrawn for research purposes.

Management Reports.—Many otherwise good testing programs tend to stagnate or become lost because top management is not kept informed. It should be the policy of those in charge

of testing to provide periodic progress reports for this purpose. A few simple rules should be followed:

1. The report should be brief, no more than two or three pages when possible.
2. It should stress true accomplishment rather than effort. A statement of what was accomplished with respect to one job is more impressive than any statement of how many were randomly tested.
3. Graphic presentations similar to those used throughout this book and described in Chap. IV should be used.
4. Involved statistics (such as probable errors, coefficients of correlation, etc.) should usually be avoided in the report. Those statistical techniques which must be used should be kept behind the scenes except in unusual cases.

SUMMARY

A few "don'ts" adapted from Ruch[1] are in order by way of summary:

1. Don't expect one test to solve all your problems.
2. Don't rely exclusively upon the advice of professors. (This book was written by a professor.)
3. Don't expect to carry over a complete program from another concern.
4. Don't regard tests solely as a means of rejecting employees.
5. Don't expect perfection in a testing program.
6. Don't let the name of a test mislead you.
7. Don't lump subtest scores into totals; analyze them separately.
8. *Don't start a testing program until you have a capable trained person to handle it.*

[1] RUCH, FLOYD *How to use employment tests.* Bull. No. 1, Los Angeles: California Test Bureau, 1944.

APPENDIX

APPENDIX A

SAMPLING THEORY AND PRACTICE

Throughout the text proper the basic point of view has emphasized the necessity of comparing the test and job performance for two or more groups of employees. Sometimes the technique has involved the comparison of the mean or average test scores of one group that has turned out well on the job with the average test scores of another group that has not turned out well. Occasionally, the reverse procedure has been used, in which the average production of a group of high scorers was compared with the average production of a group of low scorers. In other instances comparisons between two percentages have been made. During these discussions no comment has been made regarding how large a difference must be found between the two groups before it can be taken at face value or be considered significant. The purpose of this appendix section is to provide the new worker in the field of personnel testing with procedures for evaluating such differences. Technical discussions of sampling theory are not included.

Need for Evaluating Differences.—Suppose that twenty people who take a certain preemployment test are placed on the same job. The group of twenty is divided into two, so that the ten who made the highest scores are placed in one group and the ten who made the lowest scores are placed in another. Performance records for members of the two groups are examined at the end of three months or so, and average production figures are computed. Suppose, for purposes of illustration, that these are assemblers of some kind and that the average number of units per hour assembled by the high scorers is 17.6 whereas the corresponding figure for the low scorers is 15.4. The difference obtained between the two is 2.2 units. The following questions now arise: Is this difference sufficiently large to justify the use of this particular test for the future selection of employees? Does the test really identify better assemblers, or was it just a matter of chance that the high scorers did better on the job than the low scorers? If this difference of 2.2 is considered significant, would 1.8 likewise by significant? If so, how small could the difference be and yet leave little doubt that the use of the test is justified? These questions make it evident that some technique or method must be employed to evaluate differences between means or averages and between percentages.

Differences Can Occur through Chance.—Any player of bridge, poker, or other card game knows that differences can come about through chance alone. Following a completely random shuffle, a bridge player frequently discovers that he has more cards of one suit than of the others in his hand. Although the expected number is three or four, he quite often finds five, six, seven, or even eight of one suit in his hands. Only rarely does he find more, but there are instances of the perfect bridge (all one suit) on record. If an individual were not aware of

the make-up of a standard deck of playing cards, and if he attempted to guess the make-up of the deck after having been dealt his first hand consisting of two clubs, six diamonds, three hearts, and two spades, he might make a serious error. If he were to guess that the entire deck contains three times as many diamonds as clubs, he would, of course, be wrong. Even when the odds are in favor of getting identical numbers of each suit, differences do occur through chance alone. The same is true in testing. Sometimes when two groups of employees make different average scores or come up with different average production records, we find that the differences have occurred through chance and that the groups are really equal.

DIFFERENCES IN MEANS

The Significance of Differences in Means.—Statisticians have developed ways of placing an evaluation on a given difference. They have provided ways of arriving at the amount of confidence that can be placed in a particular difference that is found. It is beyond the scope[1] of this book to discuss the statistical theory and mathematical derivation of formulas. However, below is a simple step-by-step procedure that can be applied by almost anyone.

PROCEDURE FOR EVALUATING DIFFERENCES BETWEEN MEANS OF TEST SCORES

Step 1.—Assuming that employees have been divided into groups on the basis of job performance, consider one group good (g) and the other poor (p).

Step 2.—Count the number of good employees (N_g).

Step 3.—Compute the mean (M_g) test score of the good group. (Sum the scores, and divide by N_g.)

Step 4.—Compute the variability (V_g) of the scores of the good group.

a. Square each score.
b. Add the squares.
c. Divide this sum by N_g
d. From this quantity subtract the square of the mean (M_g).

Step 5.—Repeat Steps 2, 3, and 4 using the scores for the poor group. Call these N_p, M_p, and V_p.

Step 6.—Compute the number of degrees of freedom. (Add N_g to N_p, and subtract 2.)

Step 7.—Compute the difference (D_m) between the two means. (Using M_g and M_p, subtract the smaller from the larger.)

[1] See some standard work in statistics such as LINDQUIST, E. F. *A first course in statistics.* Boston: Houghton Mifflin, 1942.

Step 8.—Compute the standard error.

a. Multiply V_g by N_g.
b. Multiply V_p by N_p.
c. Add the two products.
d. Divide this sum by "degrees of freedom." (See Step 6.)
e. Now sum N_g and N_p.
f. Multiply N_g by N_p.
g. Divide the sum obtained in line *e* by the product obtained in line *f*.
h. Multiply this quotient by the quotient obtained in *d* above.
i. Obtain the square root of this product. This final answer is the standard error of the difference between means.

Step 9.—Compute the significance ratio (t). (Divide the difference obtained in Step 7 by the estimated standard error obtained in Step 8.)

Step 10.—Using t obtained in Step 9 and the number of degrees of freedom obtained in Step 6, from Table IX determine the probability.

a. In the table locate the appropriate number of degrees of freedom in the left column.
b. On the same line find the t value equal to or next smaller than the value obtained in Step 9.
c. Note the proportion at the top of the column in which this value is found. This is the probability that the difference occurred through chance.

Meaning of Level of Confidence.—Psychologists as a general rule start their investigations with the so-called "null hypothesis" which assumes the groups to be equal or, more accurately, assumes there is no real difference and that such differences as are obtained have occurred solely through chance just as the two clubs and six diamonds occurred by chance in the bridge hand in the earlier discussion. They then proceed to test this hypothesis and to accept it or reject it with a certain level of confidence. The lower the probability of chance occurrence, the greater is the certainty of the experimenter when he rejects the null hypothesis. In other words, if he finds his t value in the 5 in 100 column, his conclusion is that the probabilities are 5 in 100 that the groups are alike (or that the obtained difference could have arisen by chance). By the same reasoning, if he finds his t value in the 1 in 100 column, he rejects the null hypothesis at the 1 per cent level, which means that there is only 1 chance in 100 (as compared with 5 in 100) that the difference could have arisen through chance alone. Generally speaking, researchers reject the null hypothesis only when the probabilities are at most 5 in 100 that the difference may be attributed to chance.

SAMPLE PROBLEM

Ice-company Employees.—On page 147, Fig. 9-20 shows the relationship between scores on the *Purdue Mechanical Adaptability Test* and the owner-manager's ratings of fourteen mechanics in an ice company. Following are the values that are plotted in that graph. (A rating of 5 is high, and 1 is low.)

TABLE IX—MINIMUM VALUES OF t (THE SIGNIFICANCE RATIO) FOR DIFFERENT DEGREES OF FREEDOM REQUIRED FOR VARIOUS LEVELS OF CONFIDENCE IN CONCLUDING THAT TEST SCORES ARE *REALLY* RELATED TO JOB SUCCESS *

Degrees of freedom (N_s+N_p-2)	Probability that obtained differences occurred through chance				
	20 in 100	5 in 100	2 in 100	1 in 100	1 in 1,000
1	3.078	12.706	31.821	63.657	636.619
2	1.886	4.303	6.965	9.925	31.598
3	1.638	3.182	4.541	5.841	12.941
4	1.533	2.776	3.747	4.604	8.610
5	1.476	2.571	3.365	4.032	6.859
6	1.440	2.447	3.143	3.707	5.959
7	1.415	2.365	2.998	3.499	5.405
8	1.397	2.306	2.896	3.355	5.041
9	1.383	2.262	2.821	3.250	4.781
10	1.372	2.228	2.764	3.169	4.587
11	1.363	2.201	2.718	3.106	4.437
12	1.356	2.179	2.681	3.055	4.318
13	1.350	2.160	2.650	3 012	4.221
14	1.345	2 145	2.624	2.977	4.140
15	1.341	2.131	2.602	2.947	4.073
16	1.337	2.120	2.583	2.921	4.015
17	1.333	2.110	2.567	2.898	3.965
18	1.330	2.101	2.552	2.878	3.922
19	1.328	2.093	2.539	2.861	3.883
20	1.325	2.086	2.528	2.845	3.850
21	1.323	2.080	2.518	2.831	3.819
22	1.321	2.074	2.508	2.819	3.792
23	1.319	2.069	2.500	2.807	3.767
24	1.318	2.064	2.492	2.797	3.745
25	1 316	2.060	2.485	2.787	3.725
26	1.315	2.056	2.479	2.779	3.707
27	1.314	2.052	2.473	2.771	3.690
28	1.313	2 048	2.467	2.763	3.674
29	1.311	2.045	2.462	2.756	3.659
30	1.310	2.042	2.457	2.750	3.646
40	1.303	2.021	2.423	2.704	3.551
60	1.296	2.000	2.390	2.660	3.460
120	1.289	1.980	2.358	2.617	3.373
∞	1.282	1.960	2.326	2.576	3.291

* Table IX is abridged from Fisher & Yates, *Statistical Tables for Biological, Agricultural, and Medical Research*, Oliver & Boyd Limited, Edinburgh, by permission of the authors and publishers.

Rating	Test Score
5	114
4	104
4	102
4	95
3	106
3	101
3	94
3	93
3	88
2	93
2	89
2	85
2	81
1	74

Using these values, the aforementioned steps have been followed below in order to determine whether or not this particular test should be used in selecting mechanics for this particular job in the future.

Step 1. — For the purposes of this analysis, mechanics rated 3, 4, or 5 were considered good and those rated 1 or 2 were considered poor.

Step 2. — $N_g = 9$

Step 3. — $M_g = 897 \div 9 = 99.7$

Step 4. — $V_g = 9{,}089.67 - 99.7^2 = 49.58$

Step 5. — $N_p = 5$
$M_p = 422 \div 5 = 84.4$
$V_p = 7{,}106.40 - 84.4^2 = 43.04$

Step 6. — Degrees of freedom = 12
$(9 + 5 - 2)$

Step 7. — $D_m = 99.7 - 84.4 = 15.3$

Step 8. — *a*. $V_g \times N_g = 446.22$
b. $V_p \times N_p = 215.20$
c. Sum = 661.42
d. Divided by 12 = 55.12
e. $N_g + N_p = 14$
f. $N_g \times N_p = 45$
g. Sum in line *e* divided by product in line *f* = 0.31
h. Quotient *g* times quotient *d* = 17.087
i. Square root = 4.13

Step 9. — $t = 15.3 \div 4.13 = 3.705$

Step 10.—Using Table IX, the degrees of freedom value of 12 is located in the left column and the values across the table on that line are examined to find the one that is the same as or next smaller than the *t* value. *Note that the value of 3.705 falls between 3.055 and* 4.318. According to the rule, 3.055 being the next smaller is selected. Since this value is in the 1 in 100 column, the conclusion is that there is only 1 chance in 100 that this difference in average scores could have occurred through chance.

In this particular problem, the *Purdue Mechanical Adaptability Test* can be used for selecting ice-company mechanics with a very high degree of assurance that it will identify those who will be considered better by the owner-manager.

DIFFERENCES IN PERCENTAGES

The Significance of Differences in Percentage.—In the same fashion as outlined in the discussion of means, percentage differences can arise through chance alone. Since many comparisons that the personnel tester makes are comparisons of percentages, a simple procedure is outlined here. Here again, extensive discussion is impossible and only a simplified step-by-step process is presented.

TABLE X

	Low criterion group	High criterion group	All
Above critical score...	*a*	*b*	*a* + *b*
Below critical score...	*c*	*d*	*c* + *d*
All....	*a* + *c*	*b* + *d*	100

TABLE XI

	Low criterion group	High criterion group	All
Above critical score...	0	43	43
Below critical score...	36	21	57
All.....	36	64	100

PROCEDURE FOR EVALUATING DIFFERENCES BETWEEN PERCENTAGES

Step 1.—Select a critical score on the test, and prepare a table like Table X, placing in space *a* the *percentage* of the whole group who were in the *low* criterion group and who were *above* the critical score. The values in cells *a*, *b*, *c*, and *d* should total 100 per cent.

Step 2.—Enter the appropriate values in the cells labeled *a* + *c*, *b* + *d*, *a* + *b*, and *c* + *d*. As indicated each of these values is the sum of the other two values in the same row or column.

Step 3.—Now multiply $a \times d$, and subtract the product from the product of $c \times b$.

Step 4.—Multiply $a + c$ by $b + d$ and also $a + b$ by $c + d$. Multiply these two products.

Step 5.—Now take the square root of the product obtained in Step 4.

Step 6.—Divide the value obtained in Step 3 by the value obtained in Step 5. (This is called the *phi coefficient.*)

Step 7.—Square the value obtained in Step 6, and multiply it by N, the total number of employees. (This is called *chi square.*)

Step 8.—Now, using Table XII, locate this value (*chi square*) or the next smaller value in the table. Note the proportion above the value. This is the probability of chance occurrence.

TABLE XII.—FOR DETERMINING PROBABILITY THAT PERCENTAGE DIFFERENCE OCCURRED THROUGH CHANCE

20 in 100	5 in 100	2 in 100	1 in 100	1 in 1,000
1.642	2.706	3.841	5.412	6.635

SAMPLE PROBLEM

Same Data.—The same data utilized above can be used for demonstrating this particular method. Referring again to the scattergram in Fig. 9-20, suppose that those rated 3, 4, or 5 are placed in the high-criterion group and that those rated 1 or 2 are placed in the low-criterion group. Suppose that a critical score of 95 is established. Table XI shows the resulting percentages. The remaining steps in the procedure for evaluating this relationship are as follows:

Step 2.—Top row, $0 + 43 = 43$
Bottom row, $36 + 21 = 57$
Left-hand column, $0 + 36 = 36$
Right-hand column, $43 + 21 = 64$

Step 3.—$(36 \times 43) - (21 \times 0) = 1{,}548$

Step 4.—$(36 \times 64) \times (43 \times 57) = 5{,}647{,}104$

Step 5.—$\sqrt{5{,}647{,}104} = 2{,}376$

Step 6.—1,548 divided by 2,376 = 0.652

Step 7.—$(0.652)^2 \times 14 = 5.952$

Step 8.—Referring to Table XII, the value next smaller than 5.952 is 5.412 which is in the 1 in 100 column. That is, the probability that such a relationship might have occurred through chance is less than 1 in 100.

Here again, there is evidence that the *Purdue Mechanical Adaptability Test* can be used effectively in the selection of ice-house mechanics in this plant.

Necessity for Statistical Analysis.—Analyses of this type *must* be made if the testing program is to succeed. Failure to use safeguards of this type have been the cause of most test-program failures. A program cannot fail when the batting odds are known, and this is the way in which they are determined.

APPENDIX B

FUNDAMENTALS OF TEST ADMINISTRATION

If testing is worth while, it is worth doing well. Test scores are nothing more or less than measures of human performance; and since human performance is variable, scores are also variable. This means that in order to make comparisons among scores every effort should be made to see that the conditions under which testing is done are the best and that they are uniform. This section is intended to help in the attainment of these goals.

The Applicant and Testing.—Although recent military service has given many applicants experience in taking tests, the fact remains that employment testing is relatively new in industry. Every applicant expects to fill out an application blank and to be interviewed, but not everyone expects to take tests. Furthermore, many applicants carry over a distinct dislike for "examinations" from the school situation to the employment situation. Particularly when the applicant feels insecure or otherwise is not confident regarding his ability, he is apt to experience a mental block. The following suggestions should be helpful:

1. Do everything possible to put the applicant at ease. Make him feel that he is not "on the spot." Explain that tests along with application blank and interview data help prevent misplacements where the employee is apt to be unhappy and unsatisfactory.
2. Avoid creating the impression that selection or rejection is solely dependent upon passing the tests. Explain that the whole picture counts.
3. Never discuss the tests ahead of time or tell him anything about what they measure.
4. Avoid the use of the term *examination*. "Employment tests," "occupational tests," or "personnel tests" are better than "tests" alone. Avoid such terms as *personality, temperament,* and *intelligence.*
5. Avoid practices with minority groups that can be construed to be discriminatory.

The Applicant and Test Results.—Whether or not the applicant should have his test performance discussed with him is a knotty problem. The author feels that there is no categorical answer but that, generally speaking, the policy should be contingent upon the personnel available for such discussion. If the company has an individual thoroughly schooled in the psychology of counseling, such a practice is acceptable and often desirable. If there is no such trained person, or if there is insufficient time for an adequate and complete interview, the task had better not be undertaken. In the event that results are discussed with

the applicant, raw scores or more complicated scores should not be employed. He should be told simply that his performance was average, above average, or below average. And finally, *he should never be allowed to feel that the tests kept him from securing a job.* Just as test scores should not be flaunted at the applicant, so should they not be broadcasted about the plant. Test scores should be considered restricted information and should be kept confidential in the same fashion as medical records.

The Testing Room.—A separate testing room should be available for best results. If a room specifically for testing is impractical, one section of a larger room should be screened off. There should be no through traffic and no curious employees or applicants watching. The noise level should be low, but complete silence is not necessary and probably not desirable. Proper lighting and ventilation should receive attention as well as appropriate temperature control; particularly in the winter when applicants are apt to be wearing heavy outer clothing the temperature should be slightly less than normal room temperature. Tables of standard height and comfortable chairs should be available. Plenty of elbow room should be provided Where several applicants are tested at a time, the arrangement should be such that all can see and be seen by the examiner. Some companies find it desirable to display a sign in the waiting room similar to the following:

NOTICE TO APPLICANTS

The Occupational Test forms you are asked to fill out for the A. B. C. Company do not determine whether you will be hired or rejected. The forms are used as part of a standard hiring procedure in order to help determine the position or job for which you are best fitted. The Occupational Tests you will fill out are not ones on which you "pass" or "fail"; they are only standardized measures of interests and abilities.

The A. B. C. Company
Personnel Department

Equipment.—The examiner should provide each applicant with two sharp pencils and should have an extra supply available. He should have two stop watches or one stop watch and a clock A wall clock is not recommended, since it tends to make some applicants nervous. The administrator should have a desk with sufficient drawers or other compartments for new and used tests. These should not be visible to the applicant. The usual precautions should be employed to prevent applicants from carrying pencils and tests away.

The keeping of a daily logbook is recommended. Daily activities should be recorded. When new editions of tests are used to replace old ones, appropriate entries should be made. Everything should be dated.

The Administrator.—The good test administrator is one who likes people and who has time for them. He looks after the applicant's physical comfort and is aware of impediments such as coats and bundles. He puts the applicant at ease, does not hurry or scold him, and generally exhibits a friendly manner. He makes a favorable impression on the applicant, has a good speaking voice, and has no offensive habits or manners. He does not smoke or chew gum while

administering tests; he dresses in a conservative fashion with no fads or outlandish styles. He avoids wearing lodge or fraternity jewelry and displays no political insignia. Whether men or women should be used as administrators cannot be answered categorically. Although it is a matter of individual personality, women if adapted to the job frequently are happier in the job than men.

Administering the Test.—The administrator should adhere religiously to the instructions provided with a particular test. Directions should be given exactly, and time limits should be observed. In administering paper-and-pencil tests the following points will be found helpful:

1. Hold test up first and explain.
2. Pass out test; again explain instructions with each applicant reading same from form.
3. Ask for questions.
4. Give practice or trial as per test instructions.
   a. Circulate and help; observe whether each is doing samples correctly.
5. Give answers to practice problems; explain if necessary.
6. Ask for questions; do not violate test with answers though.
7. Instruct on time, manner of stopping, etc.
8. Use simple language: "stop," "go," etc.
9. Help the illiterate or dull person who will fail anyway in order that he does not become embarrassed; encourage him to finish as much as possible.
   a. Any applicant who emphatically quits or gives up should be pleasantly dismissed without embarrassment to him.
   b. Be friendly; everything happens sooner or later; it's all in the day's work. Keep record of unusual happenings.
10. Guard against any possibility of accusation of discrimination.

Systematic Dismissal.—Some companies have found it profitable to hand each applicant a card similar to the following:

> The A. B. C. Company wishes to take this opportunity to thank you for the time you have taken and the interest you have shown in applying for a position with us. You have now completed the various steps of your application. The employment department will give it thorough study and consideration. You will hear from the company within two days at which time you will be told definitely whether you have been accepted or rejected for employment.
>
> Sincerely yours,
> John J. Jones
> Personnel Director
> A. B. C. Company

Obviously, whatever commitment regarding notification is made should be adhered to rigorously.

Summary.—The purpose of this section has been to touch the high spots in the whole area of administration. If these suggestions are followed, they should keep the neophyte from serious pitfalls until personal experience has been accumulated.

APPENDIX C

COMMERCIALLY AVAILABLE TESTS

Any selected bibliography of tests is sure to omit tests that have been found useful by someone in the industrial field. However, since it is impossible to present an all-inclusive list,[1] a few selected titles are presented on the following pages. Those tests which have been referred to in the text and are commercially available are listed. In addition, the list also includes other tests that have already demonstrated their applicability in industry or seem promising to the author

Following are the names and addresses of a number of suppliers and publishers of tests. Most of those listed will supply catalogues or descriptive material on request. Note that each is coded with a letter and that the code letter only is used in the list of tests.

(A) American Optical Company
70 West 40th St.
New York, N. Y.

(B) Bausch and Lomb Optical Company
635 St. Paul St.
Rochester 2, N Y.

(C) California Test Bureau
5916 Hollywood Blvd.
Los Angeles 28, Calif.

(D) Humm Personnel Service
1219 West Twelfth St.
Los Angeles 15, Calif.

(E) Keystone View Company
Meadville, Pa.

(F) The Psychological Corporation
522 Fifth Ave.
New York 18, N. Y.

(G) Public School Publishing Company
509–513 North East St.
Bloomington, Ill.

(H) Division of Applied Psychology
Purdue University
Lafayette, Ind.

[1] For a comprehensive list of tests of all kinds see Buros, Oscar K. (ed.). *The 1940 mental measurements yearbook.* Highland Park, N. J.: Mental Measurements Yearbook, 1941. 674 pp.

(I) Science Research Associates, Inc.
228 South Wabash Ave.
Chicago 4, Ill.
(J) Sheridan Supply Company
P.O. Box 837
Beverly Hills, Calif.
(K) Stanford University Press
Stanford University, Calif.
(L) Stevens Institute of Technology
Hoboken, N. J.
(M) C. H. Stoelting Company
424 North Homan Ave.
Chicago, Ill.
(N) World Book Company
Yonkers, N. Y.

While some tests are difficult to classify, the groupings set up below will be generally meaningful to the reader. Page designations indicate reference to the test in this book. The designation (0) is used for those tests that are privately printed or are not generally available.

MENTAL ABILITY AND CLASSIFICATION TESTS

Adaptability Test
Tiffin and Lawshe (I)..... 51, 52, 68–74, 171
Army General Classification Test (I)..... ..... 50, 60–61
Oral Directions Test
Langmuir (F)
Otis Self-administering Test of Mental Ability (N)..... 67, 131, 134, 135, 143, 144, 163, 168, 172
Personnel Test
Wonderlic (F)..... 68, 152, 155, 157
Purdue Industrial Training Classification Test
Lawshe and Montoux (I)..... 41–53, 129, 139, 149
Senior Classification Test
Pressey (G)..... 147, 154, 166
Senior Verifying Test
Pressey (G)..... 147, 154
SRA Verbal Classification Form
Thurstone and Thurstone (I)
SRA Non-verbal Classification Form
McMurry and Johnson (I)
Scoville Mental Ability Test (F)..... 153

MANUAL AND MANIPULATIVE TESTS

Blum Sewing Machine Test (0)..... 130
Finger Dexterity Test
O'Connor (L)..... 129, 134, 140, 154

MECHANICAL APTITUDE OR COMPREHENSION TESTS

STENOGRAPHIC AND CLERICAL TESTS

TEMPERAMENT AND PERSONALITY TESTS

INTEREST TESTS

TRADE TESTS

VISUAL SKILL TESTS

ARITHMETIC AND MATHEMATICS TESTS

VOCABULARY TESTS

MISCELLANEOUS TESTS

INDEX*

* A complete index of test titles referred to in the book appears in Appendix C, pages 217–220.